REPORT RESUMES

ED 003 880 46

EXPERIMENTAL REFINING AND STRENGTHENING OF UNDERGRADUATE FOREIGN LANGUAGE CURRICULA WITH SUPPORTING RESEARCH IN TEACHING TECHNIQUES.

BY- STEWART, BLAIR AND OTHERS
ASSOCIATED COLLEGES OF THE MIDWEST, CHICAGO, ILL.
REPORT NUMBER NDEA-VI-73 PUB DATE 64
CONTRACT OEC-SAE-8833
EDRS PRICE MF-$0.45 HC-$11.56 289P.

DESCRIPTORS- *LINGUISTICS, LANGUAGES, *LANGUAGES, *LANGUAGE INSTRUCTION, *CURRICULUM DEVELOPMENT, *LANGUAGE LABORATORIES, LANGUAGE LABORATORY EQUIPMENT, PROGRAMED INSTRUCTION, INSTRUCTIONAL MATERIALS, WORKSHOPS, SUMMER WORKSHOPS, COLLEGES, BELOIT COLLEGE, CHICAGO, ILLINOIS

A 4-YEAR PROGRAM OF COOPERATIVE ACTIVITY IN FOREIGN LANGUAGE INSTRUCTION WAS UNDERTAKEN BY 10 MIDWEST COLLEGES. THE BASIC OBJECTIVE OF THE PROGRAM WAS TO IMPROVE TEACHING AND CURRICULAR PRACTICES FOR FOREIGN LANGUAGE MAJORS AND FOREIGN LANGUAGE TEACHING MAJORS. UNDER THE DIRECTION OF A FULL-TIME COORDINATOR, AN OFFICE FOR THE PROGRAM WAS ESTABLISHED AT BELOIT COLLEGE. PART-TIME COORDINATORS WERE APPOINTED AT EACH PARTICIPATING COLLEGE. FACULTY WORKSHOPS WERE HELD EACH SUMMER FOR COORDINATING AND IMPLEMENTING THE PROGRAM. A TOTAL OF 52 FACULTY MEMBERS ATTENDED THE SUMMER SESSIONS WHICH INVOLVED REVIEWS OF LANGUAGE LABORATORY EQUIPMENT, THE DESIGN OF CLASSROOM EXPERIMENTS, STUDY OF THE RELATION OF LINGUISTICS TO LANGUAGE INSTRUCTION, AND THE DEVELOPMENT OF INSTRUCTIONAL MATERIALS. THE WORKSHOPS WERE RATED BY THE PARTICIPANTS AS THE MOST VALUABLE ASPECT OF THE PROGRAM. A NUMBER OF TENTATIVE CONCLUSIONS FROM THE EXPERIMENTAL PHASES OF THE PROGRAM WERE PRESENTED AND DISCUSSED. (JC)

73

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

EXPERIMENTAL REFINING AND STRENGTHENING OF UNDERGRADUATE FOREIGN LANGUAGE CURRICULA WITH SUPPORTING RESEARCH IN TEACHING TECHNIQUES

FINAL REPORT
O.E. CONTRACT NO. 8933

TO: Director, Language Development Branch
Division of College and University Assistance
United States Office of Education
Department of Health, Education, and Welfare

FROM: THE ASSOCIATED COLLEGES OF THE MIDWEST

THE RESEARCH REPORTED HEREIN WAS PERFORMED PURSUANT TO A CONTRACT WITH THE U.S. OFFICE OF EDUCATION, DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, NDEA.

O.E. CONTRACT SAE 8933

TERM: 30 MAY 1960 TO 14 JUNE 1964

ED 003880

PRECEDING PAGE BLANK-NOT FILMED

FOREWORD

This is a report of the results of U.S.O.E. Contract No. SAE 8833. The original contract covered the period, June 15, 1960 to June 14, 1963. It was extended by a supplementary contract to June 14, 1964.

The report has been the work of several hands Klaus A. Mueller, Coordinator of Language Instruction Programs of the Associated Colleges of the Midwest, and William Wiersma, Consultant on Psychometrics, submitted basic materials. The final draft was the work of Blair Stewart, Joseph Stokes, and E. K. Williams. The chief editorial responsibility has been carried by Blair Stewart.

In presenting information with respect to activities at the member colleges of the Associated Colleges of the Midwest, the colleges have been identified by name except when intercollege comparisons of student performance are involved. In these instances Roman numerals have been used. The numerals assigned to colleges vary from one part of the report to another, but are consistent within a given context.

The consultants that assisted in this program are named, and their roles described, in the first chapter.

The members of the language departments at the member colleges who for various terms served as local coordinators were:

Beloit College:	John F. Hartman, Donald A. Murray
Coe College:	Herbert F. Wiese
Cornell College:	F. Alan DuVal
Grinnell College:	F. Andrew Brown, Dirk Baay
Knox College:	Miriam E. Wilson, Elna Jeffries, Sherman Brown
Lawrence College:	Richard Winslow, Dorrit Friedlander
Monmouth College:	Mary Eleanor Maule, Dorothy Donald, Johann Struth

Ripon College: Daniel F. Delakas,
Henry Schneider,
Alexander Hooker

St. Olaf College: Loring Knecht

PRECEDING PAGE BLANK-NOT FILMED

TABLE OF CONTENTS

Chapter Page

Foreword ii

I. Purposes and Procedures 1

II. The Audio-lingual Emphasis 11

III. The Testing Program 40

IV. The Language Laboratory 81

V. Individualization of Instruction: Placement 104

VI. Individualization of Instruction and the Language Laboratory 132

VII. Individualization of Instruction: Programmed Learning 156

VIII. Pace and Mastery 207

IX. Attitudes and Motivation 215

X. Conclusion 268

Appendices

Appendix I: ACM Speaking Tests: 276

1. First Semester 276

2. First Draft, Second Semester 1961-62 289

3. Revised Drafts, Second Semester 1961-62 299

Appendix II: Pre Test Scores and Elementary Level Post-test Scores by College 328

Appendix III: Questionnaires to Students and Faculty on Programmed Learning 341

Appendix IV. Psychometric Tests:

1. Sample Psychology Instruments 353

2. Analysis of Data from Psychological Tests 383

PRECEDING PAGE BLANK NOT FILMED

LIST OF TABLES

Number Page

1. Lengths of Withholding Period and Number of Sections Involved in Colleges Participating in Withholding Orthography Experiment 27

2. Average Test Scores and Course Grades of Experimental and Control Sections in the Same College: Experiment in Withholding Orthography 32

3. Average Test Scores for Two Languages for Colleges Withholding Orthography Compared with No Withholding 34

4. Average Test Scores for All Colleges Participating: Orthography Withheld compared with No Withholding 36

5. Method of Administering ACM Speaking Tests in January, 1962 52

6. General Summary of Test Results 54

7. Relative Standing of Males and Females on Pre Tests and Post-tests for Groups Consisting of Those Tested at Each College in Each Language 55

8. Correlation Matrix for French Students, First Semester 1961-62 57

9. Correlation Matrix for German Students, First Semester 1961-62 59

10. Correlation Matrix for Spanish Students, First Semester, 1961-62 61

11. Correlation Matrix for French Students, End of 1962-63 65

12. Correlation Matrix for German Students, End of 1962-63 67

LIST OF TABLES
(continued)

Number		Page
13.	Correlation Matrix for Spanish Students, End of 1962-63	70
14.	Mean Scores on Modern Language Association Proficiency Tests	73
15.	Language Laboratory Installations at Member Colleges	84
16.	Results of Experimentation with Laboratory and Non-Laboratory Sections of Elementary French at St. Olaf College, 1961-62	90
17.	Average Test Scores and Course Grades by Language for Groups Using Language Laboratories Compared with Those Not Using Laboratories	93
18.	Average Test Scores and Course Grades by Language for All Colleges Participating in Comparison of Language Laboratory Situation with No Laboratory	95
19.	Average Achievement Scores and Course Grades by Language for All Colleges Participating: Comparison of Different Types of Laboratory Installation, First Semester 1961-62	100
20.	Average Achievement Scores and Course Grades by Language for All Colleges Participating: Comparison of Different Types of Laboratory Installation, Second Semester 1961-62	101
21.	Average Achievement Scores and Course Grades by Language for All Colleges Participating: Comparison of Different Types of Laboratory Installation, 1962-63	103
22.	Results of Placement Process at Five Colleges: All Languages	122
23.	Percentage of Students Placed at Equivalence Levels by Language and by College	126

LIST OF TABLES
(continued)

Number Page

24. Percentage of Students Placed at Equivalence or More Advanced Level, by Language, Selected Colleges, 1959-63 127

25. Post-test Results and Course Grades for Monitored and Unmonitored Laboratories 147

26. Post-test Scores and Course Grades: Take-Home Discs Compared with Non-Monitored Laboratory 154

27. Average Achievement Scores and Course Grades by Language and College: Comparison between Programmed Learning and Conventional Materials, 1962-63 165

28. Average Achievement Scores and Course Grades by Language and College: Comparison between Experimental and Conventional Materials, 1962-63 166

29. Average Achievement Scores and Course Grades in German for All Colleges Participating: Comparison between Programmed Learning and Experimental Materials 168

30. Distribution of Responses to the Comparison Questions by Area of Mastery and Rating 172

31. Distribution of Comparisons of Programmed versus Non-programmed Materials by Effectiveness and Student Aptitude 174

32. Distribution of Student Responses by Number and Percentages to Programmed Learning 181

33. Comparison of Student Attitudes Toward Programmed Learning as Compared with More Conventional Instructional Methods 183

34. Distribution of Responses of Students Preferring Programmed Learning by Category and Language 186

LIST OF TABLES
(continued)

Number		Page
35.	Mean Ratings of Programmed Learning by German Students, by College and Item	188
36.	Mean Ratings of Programmed Learning by Spanish Students by College and Item	189
37.	Frequency, Mean, Median and Range of Number of Weeks Required to Complete Programmed Materials by College and Language	192
38.	Frequency, Mean, Median, and Range of Number of Weeks Required to Complete Programmed Materials by College and Language	193
39.	Distribution of Responses of Students as to Why Programmed Learning Materials were Easier than Conventional Materials, by Category and Language	196
40.	Distribution of Responses to Question on the Absence of Explanations of Grammar: by Response and Language	199
41.	Distribution of Student Feeling About Group Meetings, by Language and Response	201
42.	Distribution of Useful Features of the Group Meetings, by Category and Language	202
43.	Mean Scores on Psychological Tests by College	229
44.	Correlations Between Attitude Measures; German Control Group; June 1963	235
45.	Intercorrelations of Pre-Test and Achievement Measures; German Control Group; June 1963	236
46.	Means, F-Ratios, and Frequencies for the Psychometric Tests; Withholding Orthography Experiment; June 1963	238
47.	Means, F-Ratios, and Frequencies for the Psychometric Tests; Monitoring Experiment; June, 1963	240

LIST OF TABLES
(continued)

Number		Page
48	Means, F-Ratios, and Frequencies for Psychometric Tests; Laboratory Type Experiment, June 1963	241
49	Means, F-Ratios, and Frequencies for Psychometric Tests; Programmed Learning Experiment; June 1963	244
50	Mean Scores and Total Frequency for the Pre-College Control Group on the Eleven Psychometric Tests; 1963	245
51	Relative Positions of Means by Group and Test; Withholding Orthography and Monitoring Experiments; 1963	246
52	Relative Positions of Means by Group and Test; Laboratory Type and Programmed Learning Experiments, 1963	247
53	Distribution of "Integrative-Instrumental" Motivation Scores, Test I-A	253
54	Distribution of "Integrative-Instrumental" Motivation Scores, Test I-B	255
55	"Thinking in the Language" Responses to Question:	259
56	"Thinking in the Language" Responses to Question:	260
57	"Thinking in the Language" Responses to Question:	262
58	"Thinking in the Language" Responses to Question:	263
59	"Thinking in the Language: Responses to Question:	264
60	"Thinking in the Language" Responses to Question:	266

LIST OF CHARTS

Chart		Page
1	Mean Scores on Modern Language Association Proficiency Tests: French	74
2.	Mean Scores on Modern Language Association Proficiency Tests: German	77
3.	Mean Scores on Modern Language Association Proficiency Tests: Spanish	78
4.	Percentage of Students Placed at Equivalence Level, Selected Colleges, 1959-1963	128

I Purposes and Procedures

Purposes

Instruction in foreign languages and literatures is an important aspect of the liberal arts curriculum. Colleges of liberal arts are therefore concerned with criticisms of foreign language instruction and with proposals for change. The supply of such criticisms and proposals has been very large in recent years but the relative validity of the various criticisms and the value of the proposals are generally undetermined. As a consequence, when ten midwestern liberal arts colleges joined together in an attempt "to increase their educational effectiveness and operating efficiency" a program of cooperative activity in the field of language instruction was adopted. It was felt that joint consideration by ten colleges of the aims of foreign language instruction and careful examination and testing of proposals for improvement would provide assistance to the individual colleges in their search for ways to improve the instructional process.

In carrying out this project a contract was entered into with the United States Office of Education. This contract was for a three-year period from June 15, 1960 to June 14, 1963. Subsequently a supplemental contract was signed covering the period from June 15, 1963 to June 14, 1964. This is a report of the work done under the original and supplemental contracts.

The focus of the work under the contract was the "improvement of instruction in foreign languages and literatures at all levels, including the pre-service, undergraduate course of training for prospective foreign language teachers." A more explicit statement in the contract reads, "The basic objective of the project is to improve teaching and curricular practices throughout the undergraduate years to the end that foreign language majors and foreign language teaching majors may normally be expected to achieve at least the 'minimal' rating on applicable parts of the MLA-Starr proficiency tests."

The "MLA-Starr proficiency tests," now known as the Modern Language Association Foreign Language Proficiency Tests for Teachers and Advanced Students provide for the measurement of the abilities specified in the statement on "Qualifications for Secondary School Teachers of Modern Foreign Languages" issued in February 1955 by the Modern Language Association and later endorsed by eighteen national and regional organizations of language teachers. The seven competencies defined in the statement are: 1) listening comprehension, 2) speaking, 3) reading, 4) writing, 5) language analysis, 6) culture, and 7) professional preparation.

In the effort to strengthen foreign language instruction at all levels, member colleges in the Associated Colleges of the Midwest (ACM) were to use and explore a number of generally accepted practices, to make use of standardized achievement tests, and to conduct experiments to test the effectiveness of different procedures. The practices included the use of foreign language placement tests, the use of electromechanical facilities to increase the amount of audio-lingual practice, provision of facilities for and encouragement of self-directed study, the use of lay native speakers to assist the regular faculty, recognition of student progress on the basis of demonstrated proficiency rather than exclusively on a credit hour basis, the near-exclusive use of the foreign language in all upper-level courses in literature and languages, the identification and use of optimum teaching methods and the coordination of language instruction.

When the contract was signed it was expected that two major new foreign languages tests would be available for use in cooperative testing programs at the member colleges. These tests were the National Defense Education Act-supported Modern Language Association Cooperative Classroom Achievement Tests and the Foreign Language Proficiency Tests referred to above. In a supplemental agreement the testing program was expanded by including attitudinal and motivational measures used in the study, Extended Classroom Experimentation with Varied Sequencing of the Four Skills in German Instruction, made at the University of Colorado under a contract with the United States Office of Education and directed by George A.C. Sherer. The tests used were prepared by Michael Wertheimer and were based on measures first developed by Wallace E. Lambert.

The third element in the program was a series of classroom experiments designed to shed light on instructional procedures. The problems listed in the initial formulation included: the optimum length of the foreign language laboratory periods and the optimum distribution of the students' time between class meetings, laboratory and homework; the best type of language laboratories; the optimum degree of mastery which should be required before the student proceeds to new material; the feasibility and efficiency of a learning procedure in which each student proceeds entirely at his own pace, and the relative value of different types of stimuli -- aural, visual and kinesthetic.

The purposes of the program set forth above add up to a large and varied list. Although changes were made in the study as it progressed, the main features and the broad purposes envisaged at the outset were preserved. As would be expected in a study of this type, some of the subsidiary purposes were modified, some were abandoned, and some new purposes were added.

Procedures

The program was under the direction of a full-time coordinator, Klaus A. Mueller. He established an office at Beloit College and made frequent trips to the other colleges. Part-time coordinators were appointed at each college to take the responsibility for communication and reporting between the colleges and the Office of the Language Coordinator at Beloit. Normally the local coordinators were released from one-third of their teaching responsibilities to provide time for these activities. Contact with the local coordinators was maintained by letters, memoranda and personal visits by the Language Coordinator. Detailed policies were developed by the Summer Faculty Workshops, attended by the local coordinators or language department chairmen, and other members of the language departments of member colleges. As participants in the workshops, the language coordinators were responsible for transmitting knowledge of policies developed to their colleagues and for collecting information and suggestions for submission to the Office of the Coordinator. During the course of the program 15 News Bulletins were issued by the Coordinator and distributed to all language faculty members through the local coordinators. Supplements to the News Bulletins disseminated bibliographic materials on language texts and language teaching, linguistics and teaching aids.

The Faculty Workshops held each summer were a major device for coordination and for implementing the program. The purposes of the workshops broadly stated were: to consider the relation of linguistics to language instruction, to review the types of language laboratory equipment available, to explore and develop materials for language instruction in classroom, laboratory and at home, to provide direct contact with audio-lingual methods of instruction, to design classroom experiments and review experimental results, and to plan future activities in materials production, testing and experimentation.

The consideration of the relation of linguistics to language instruction was promoted by a series of lectures at each workshop and the circulation of reading materials. Lectures were given by Klaus A. Mueller, by Val Hempel and Boris Jordan of the U. S. Army Language School, and by William E. Bull of the University of California at Los Angeles. A number of films on linguistics and audio-lingual techniques were shown as well as film strips and slides demonstrating audio-lingual materials.

The study of laboratory equipment at the Summer Faculty Workshops was promoted by exhibits, demonstrations and group discussions. Group meetings were also important in the field of materials production. There were lectures by William E. Bull on the construction and use of his "Visual Grammar of Spanish." John B. Carroll of Harvard University, Marcel Moraud of Hamilton College, Stanley M. Sapon of the University of Rochester and Maurice W. Sullivan, then of Hollins College, and Dale Brethower of the University of Michigan, an associate of Rand Morton, gave lectures and participated in discussions of programmed learning of foreign languages.

Direct contact with audio-lingual instruction was achieved by having the members of the workshops study an unfamiliar language taught by a person long-experienced in the audio-lingual approach. In 1960 and 1961 they studied Danish, instructed by Val Hempel. In 1962 they were taught Bulgarian by Boris Jordan. These teachers also lectured on the audio-lingual approach as did Alfred S. Hayes, Foreign Language Consultant, and Leroy James Benoit, then Chief of the Area and Language School Training Division of the U. S. Information Agency.

The activities connected with testing and experimentation were carried out with some assistance from visiting lecturers, but mostly in group discussions, frequently assisted by consultants. Assistance of this kind was provided by John B. Carroll of Harvard University, Reemt R. Baumann of the University of Wisconsin, Donald J. Malcolm of the Educational Testing Service, Paul Pimsleur of Ohio State University and William Wiersma of the University of Wisconsin, now at the University of Toledo.

Much of the effective work done in the Summer Workshops was achieved in informal working and discussion groups. The participants lived in Beloit College dormitories and had their meals together in the college commons. Total enrollment for the three years was 76, an average of approximately two and a half persons per college per year. For individual colleges the number of enrollments ranged from 4 to 13 or annual averages ranging from slightly more than 1 to more than 4 per year.

The substantive activities of the program in curricular development, experimentation and testing were carried out at the individual colleges. The local coordinators had responsibilities for communication, but major developments called for the participation of many persons. The Summer Workshops were an important means for involving faculty members in the program. The activities at the member colleges in connection with the program involved many members of the language faculties, some of whom were unable to attend the workshops. Under ACM procedures each college decides whether, and to what extent, it will participate in an ACM activity. In a program such as this, decisions on participation are also made on a departmental, or even an individual faculty member, basis. Such participation was extensive. Fifty-two different faculty members attended the Summer Workshops; 16 attended two sessions and four attended all three sessions. A very high proportion of the foreign language teachers in the ten colleges were involved in some aspect of the program.

II The Audio-Lingual Emphasis

The Objectives of Language Study

Modern foreign language teachers have been commendably concerned with the formulation as precisely as possible of the aims of foreign language teaching at the college level--as well as at other levels. It has been suggested that this is a defensive reaction arising from some sense of insecurity with respect tothe position of foreign language study in institutions of higher education. Whether this is true or not is of little moment; the essential fact is that college language teachers have stated their objectives with a forthrightness that might well be emulated by teachers in other fields.

As early as 1898, for example, a report on the objectives of language study was made by the Committee of Twelve of the Modern Language Association of America. A later study resulted in the publication in 1930 of The Teaching of Modern Foreign Languages in the United States: A Report Prepared for the Modern Foreign Language Study, written by Professor Algernon Coleman of the University of Chicago. Coleman says of the Committee of Twelve, "In their judgment the primary or 'immediate' objective was reading ability. Speaking ability they put expressly in the background."[1] Although audio-lingual skills were not ignored, the prevailing emphasis of statements of other groups concerned with the objectives of language study was on reading. All language skills were recognized as important in the report written by Coleman. He was, however, much impressed by the fact that most persons studying a foreign language in this country spent no more than two academic years in such study. If language study is to be so limited, Coleman concluded that a limited objective was desirable. The following paragraph appears in the "Summarizing Conclusions" of his chapter on the "Content of the Modern Language Course."[2]

> Since reading ability is the one objective on which all agree, classroom efforts during the first two years should center primarily on developing the ability to understand the foreign language readily through the eye and

---

[1] Algernon Coleman, The Teaching of Modern Foreign Languages in the United States: MacMillan, New York, 1930, p.7

[2] *Ibid*., p. 170

> through the ear. The goal must be to read the foreign language directly with a degree of understanding comparable to that possessed in reading the vernacular. In order that students may attain this goal, reading experiences must be adequate and the results of all other types of class exercise must converge toward the same end."

These discussions related to language study at the elementary level, both in school and college. To the pragmatic argument advanced by Coleman for concentrating on reading, others have added considerations particularly relevant to elementary language instruction at the college level. It has been pointed out that listening and speaking, as psychomotor skills, do not as such contain an intellectual component sufficient to justify inclusion in the college curriculum. This charge could not be brought against grammar, which is based on analysis of language, and a rapidly developed reading skill could be the basis for the study in the original of literary masterpieces of high intellectual and aesthetic merit.

For many years foreign language study in American colleges and universities was primarily geared to the development of reading facility as rapidly and as effectively as possible. Teaching aids were predominantly grammars, graded readings and bi-lingual dictionaries. There were, of course, exponents -- and practioners -- of the "direct method", but the major emphasis in foreign language study in this country up to and long beyond the Second World War was on reading ability supported by understanding of basic grammatical principles.

In the post-war period the emphasis shifted heavily in the direction of listening comprehension and speaking ability. This shift was accompanied, and reinforced, by the developing study of linguistics with its characterization of language as "patterned human social behavior" and the recognition that the speaking and listening aspect of this behavior far outweighed in practice and human significance the writing and reading aspect.

The definitive post-war declaration of the objectives of language study appeared as a statement on "Qualifications for Secondary School Teachers of Modern Foreign Languages" referred to in the previous chapter. The seven competencies listed in that statement included the four basic skills of listening, speaking, reading and writing. A teacher of foreign languages should have these basic skills but should also have knowledge and understanding of the culture of the countries where the language is spoken, of linguistic analysis and of methods of teaching foreign languages.

Although originally addressed to the qualifications of secondary school language teachers, this statement quickly came to be looked upon as an expression of a prevailing view of the purposes of foreign language instruction at all levels. All four of the basic skills are important and all but the seventh are accepted as appropriate general objectives of language study. It is also clear that the order in which the basic skills are listed (listening, speaking, reading, writing) is neither capricious nor accidental.

A number of other forces came to bear on the nature of foreign language study at the college level: the growing importance of the United States in international affairs and the increasing flow of Americans abroad for both long and short term stays, the audio-lingual emphasis of language instruction in the armed forces and for persons scheduled for foreign assignments, and the widespread activities of the United States Office of Education under the National Defense Education Act. The result has been a pressure which has been felt on every campus in America, a pressure to consider the extent to which the emphasis in language instruction should be changed in the direction of an increased emphasis on the spoken word.

Another development has complicated the life of the contemporary teacher of foreign languages. This is the improvement and proliferation of devices for reproducing sound. The "language laboratory" dates back at least to the Edison dictating machine with its shavable wax cylinders, but the great burgeoning of commercially promoted language laboratories came only with the widespread use of magnetic tape. Very few language teachers have remained untouched by this development. Even those who feel strongly that reading proficiency should be the primary aim of college language instruction have become interested in what these devices can contribute to the less important aims, or how they may be used by students who are interested, on their own initiative, in developing audio-lingual skills.

The ACM Program Emphasis

This is the general setting in which the present study developed. When faculty members from the ten member colleges came together for the first Faculty Summer Workshop in 1960, attention was directed particularly at the instructional methods advocated by proponents of the audio-lingual approach. It was evident that there were many firmly held, but conflicting opinions, and consideration was given to the question whether classroom experimentation might help to resolve some of the conflicts. Attempts were made to devise some experiments which might be carried out in a preliminary way during the next academic year, 1960-61, and a number of such pilot experiments were agreed upon and preliminary plans for their conduct were formulated.

Another activity that turned out to be the most generally commended feature of the annual faculty workshops was the study by the participants of an unfamiliar foreign language taught by strictly audio-lingual methods. College foreign language teachers whose earlier formal study of foreign languages had in many cases been of a more traditional sort, had for the first time the experience of a complete novice being exposed to language instruction entirely divorced from reading and writing. Most of them gained a realization of the potentialities as well as the difficulties and limitations of such an approach. Some of them found the experience invaluable as they strove to increase the audio-lingual emphasis in their own approach to teaching.

One of the basic questions discussed was that of the possible sacrifice in reading competence involved in the acceptance of an audio-lingual method. Some enthusiasts for the latter approach insisted that within the time span of normal college language study students by this method would achieve as much, if not greater, proficiency in reading as students taught by methods which emphasized reading. If this is true, if it is possible to "have all this and heaven too," surely no one could object to the adoption of the audio-lingual approach. Is it possible to discover whether this is true? Would it be feasible to set up an experiment with matched groups of students studying the same structural and lexical items and with the same instructor, and measure the results in all four skills at the end of one year of college study, and, possibly, at the end of two years?

Participants in the first workshop were aware of the details of an experiment of exactly this sort just then being initiated under an Office of Education contract at the University of Colorado. Contemplation of the problems of experimental design and conduct in the ACM program revealed that it was not feasible, at least for these colleges at this time. Each method is most effectively used with materials particularly suited to it. It would have been necessary, as at Colorado, to prepare two complete sets of materials covering the same vocabulary and structural items, but adapted to the two different instructional methods. The lack of staff with the time to prepare such materials and difficulties in controlling instruction and other variables led to the conclusions that such a direct contest was not feasible.

There were a number of other less global experiments, all of them involving aspects of the audio-lingual approach which could be attempted. If this were combined with a uniform testing program information should emerge as to the results of the audio-lingual approach to the acquisition of the four basic language skills. The fortress might not be taken by direct assault, but it might still fall as the result of a large number of sapping operations each undermining a small section of the edifice.

The Withholding of Orthography

One such experiment had to do with the withholding of orthography in the early stages of language learning. It is a widely held tenet of the audio-lingual school that the initial phases of learning a foreign language should involve the spoken language only, with no reference to the printed or written forms. It is pointed out that we all learned our native tongues in this manner. The method has been successfully used in language instruction at the Army Language School and elsewhere, and it is widely adopted as the appropriate procedure for language instruction in the elementary grades. Furthermore, instructors who have used it have reported a deterioration in pronunciation when orthography is introduced -- although presumably not to the level that would prevail if orthography had been introduced at the outset!

The arguments for beginning language instruction by entirely audio-linqual means, without giving the student any written material, can be summarized in a few simple statements. Our ability to speak English is made up of a set of habits in forming certain sound patterns ourselves and recognizing the equivalent patterns when made by others. A college student for obvious reasons cannot learn a second language as he did his English through the slow, natural processes of early childhood, but he needs the closest feasible approximation of the detailed practice, correction, and renewed practice which the child has. Such reiteration requires time, and more time is available to the teacher and the student when the drill is oral rather than written. Furthermore, the learning process is more effective in its impact because of the direct personal contact.

The exponents of the audio-lingual approach hold that the speaking and listening processes are primary, in that they embody the living corpus of a language, while the reading and writing aspects are more peripheral. Certainly the natural process of learning English is to proceed from a fair mastery of the spoken language to later learning of how to read and write; so the audio-lingual approach seems to follow a natural sequence. Also the fact that language is essentially communication is more evident and influential in audio-lingual instruction.

The association deeply ingrained in the speaker of English between sounds, words, and expressions and their written symbols induces a very strong resistance to the association of the same symbols with different sounds. Therefore, the introduction of orthography too soon in the student's first contact with a new language will tend to slow down his accommodation to the sounds of that new language because he associates the written characters so compulsively with other sound values.

At the Summer Faculty Workshop in 1960 considerable attention was given to the possibilities of classroom experimentation in foreign language teaching techniques.

Professor John Carroll addressed the workshop and talked with groups concerned with particular problems. It was clear that the skills, the time and the facilities were not available to make it possible to design and conduct during the academic year 1960-61 experiments that were sufficiently rigorous to warrant that designation. On the other hand, members of the workshop felt that mere innovation was not enough. The changes to be tried would -- as with any innovation -- be subjectively appraised. But attempts were also to be made to introduce some experimental features; experimental and control groups would be established, attempts would be made to find and administer suitable pre-tests, experimental conditions would be specified as precisely as possible and uniform post-tests would be administered. It was recognized that performance in most of these respects would be inadequate, and the experimental activities of the first year should be considered to be exploratory in nature.

Experiments with the withholding of orthography during the first weeks of elementary language instruction were carried out at four colleges in 1960-61; two languages were involved at one college and one language only at each of the other three. Twelve sections and eight different instructors participated. The students had all taken the College Entrance Examination Board Scholastic Aptitude Test and, at the beginning of the year they were all given the Modern Language Aptitude Test for pre-test purposes.

The experimental sections started their study of the language with no reference to the written or printed form of the language being studied for periods ranging from one to nine weeks. Classes were devoted to speaking and listening to the language. Various props and cues were used, such as objects in the classroom, pictures, stick figures and English words. Homework involved listening and speaking the language with the aid of magnetic tapes and phonograph records. Language laboratories had been installed at two of the colleges; installation was under way at a third and had not begun at the fourth. Practically all of the homework was done in the laboratories at the first two colleges. At the other colleges students met in groups to listen and speak in unison with the assistance of tape recorders or phonographs.

Problems in Experimentation

There had been little time for the preparation of special materials for use during the initial period. The instructors did what they could to prepare or obtain teaching aids suitable for purely audio-lingual instruction. Eight of the sections were in Spanish and the text used was Modern Spanish. This book was the result of a cooperative effort sponsored by the Modern Language Association to prepare a textbook with accompanying phonograph records and magnetic tapes to provide a year of instruction using audio-lingual methods. Each section is based on a recorded conversation which students are expected to memorize. There are a large number of drills, recorded on tape. These materials were very useful, but they were not designed for use without the text. In French and German the problems of suitable materials were even more difficult.

It was initially proposed that in about half of the experimental sections orthography should be withheld for three weeks and in the other half for six weeks. When instruction actually got under way most of the instructors found it difficult to adhere to the original plans. The number of sections in which orthography was withheld and the lengths of time of withholding periods are shown in Table 1. There was a range of from five class meetings to nine weeks in the length of the withholding period, with five sections at the minimum and one at the maximum. Of the 26 experimental sections, five witheld for five class sessions, nine withheld for two weeks, one for two and a half, six for three weeks, three for four weeks and one each for seven and a half and for nine weeks. In all, seven colleges were involved, but only one participated in all three years. One college participated for two years and the remaining five for one year only.

The major factor in the departure from the original experimental plan and the brief participation of most of the colleges was the subjective appraisal of class progress by the instructors. These experiments were an attempt to adapt purely audio-lingual techniques to college groups ranging in numbers from 15 to over 30. These techniques had previously been used in a number of settings, including large groups at the elementary school level and adult instruction in groups of two or three, frequently with the assistance of a

TABLE 1

LENGTHS OF WITHHOLDING PERIOD AND NUMBER OF SECTIONS INVOLVED IN COLLEGES PARTICIPATING IN WITHHOLDING ORTHOGRAPHY EXPERIMENT

Year and College	French Sections No.	French Withhold Weeks	German Sections No.	German Withheld Weeks	Spanish Sections No.	Spanish Withheld Weeks
1960-61						
College A	2	2(1)	-	-	3	2
College B	-	-	-	-	2	3
College C	-	-	-	-	3	4
College D	-	-	1	9	-	-
1961-62						
College D	5	1+(2)	-	-	-	-
College E	-	-	-	-	2	2
College F	1	2	1	2½	1	3
1962-63						
College D	-	-	-	-	1	2(3)
College F	1	3	1	3	1	3
College G	1	7½	-	-	-	-

(1) With a third week using the text, but with no writing.

(2) Actually five class meetings.

(3) Actually eight class meetings.

large amount of special materials and free access to tape recorders and phonographs. One of the greatest difficulties encountered in the attempt to adapt this approach for college classroom use was the provision of sufficient materials, facilities and motivation to induce the student to devote as much time to study of the language as he would have spent with more conventional procedures.

Another problem was the phenomenon which from the experimental point of view would be called "contamination." College students are literate and they are accustomed to studying books. They frequently feel frustrated when denied access to text materials. The case was particularly clear-cut where the Modern Language Association Spanish text was being used and there were experimental and control groups taught by the same teacher. Students in the experimental groups must frequently have had a lively curiosity about the contents of the textbooks belonging to their fellow-students in the control sections!

The withholding experiment was the result of concern with respect to the lingual aspect of the audio-lingual emphasis. The aim was to improve pronunciation and other aspects of speaking. In the first year use was made of the listening comprehension part of the College Entrance Examination Board's Achievement Tests. The assumption was that the two audio-lingual skills are highly correlated, and that listening comprehension would be an adequate measure of speaking ability. This was a highly uncomfortable assumption -- to be dropped at the earliest possible moment.

To meet this problem time was devoted during the 1961 Faculty Workshop to the development of tests of speaking ability to be used in 1961-62. These tests are more fully discussed in the chapter on the testing program. They were administered at the end of each semester in 1961-62.

In 1962-63 the MLA Cooperative Foreign Language Tests became available in preliminary form. These tests measure all four of the basic language skills. They were administered at the end of the academic year to the students participating in the experimental program and were centrally scored by the Educational Testing Service.

The Results of Experimentation

The problem of "contamination" has been mentioned above. Because of this problem, and for other reasons, two types of experimental design were used in the withholding study. One type involved experimental and control sections in the same college. In the other all the elementary sections in a given language at the college were subject to withholding. This reduced the danger of contamination, but it made it necessary to compare results between colleges to get some basis for judging the results of withholding. Although this procedure has its limitations, meaningful comparisons are possible. The withholding experiment was aimed at speaking. It is, therefore, relevant to ask the relative standings of the students in the experimenting colleges on the speaking tests as compared with their standings on the other skills.

In all three years there were withholding experiments in German with an experimental and a control section paired. In French there were two colleges where comparisons of this type were possible. Only one college had paired sections in Spanish for one year, 1962-63. The results of these experiments in scores that have been adjusted on the basis of the SAT and MLAT pretests are given in Table 2. Also shown in the table is the estimate of statistical significance, the probability that differences of the order shown might arise by chance. The most significant level indicated, 0.01, represents a probability of one in a hundred; 0.05 represents a probability of five in a hundred. Probabilities greater than 0.05 are indicated by N.S. -- Not Significant.

Study of Table 2 will reveal few statistically significant differences between the experimental and control groups. Withholding seems to have contributed to greater listening ability in German at College II in 1961-62, and to reading in Spanish at the same college in 1962-63. But in reading in German at College I in 1960-61 and in both speaking and listening in French at College II in 1962-63 the students in the control groups did better than those from whom orthography was withheld at the beginning of the course.

When comparisons are made between results at a college where orthography was withheld and those at other colleges

TABLE 2

AVERAGE[1] TEST SCORES AND COURSE GRADES[2] OF EXPERIMENTAL AND CONTROL SECTIONS IN THE SAME COLLEGE EXPERIMENT IN WITHHOLDING ORTHOGRAPHY

Year and Language	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
1960-61										
German (College I)										
Withholding (9 weeks)	-	-	20	14	20	43	-	-	-	-
No Withholding	-	-	20	14	20	46	-	-	-	-
Significance				NS		.05				
1961-62										
German (College II)										
Withholding(2-1/2 weeks)	-	-	18	48	18	44	-	-	18	4.0
No Withholding	-	-	14	44	14	44	-	-	15	3.9
Significance				.01		NS				NS
1962-63										
Spanish (College II)										
Withholding (3 weeks)	18	58	18	23	19	27	18	84	19	4.0
No Withholding	16	50	18	19	18	18	3	57	18	3.4
Significance		NS		NS		.05		No Test		.05
French (College II)										
Withholding (3 weeks)	20	49	21	25	27	23	-	-	27	3.4
No Withholding	53	55	59	30	60	37	-	-	60	3.3
Significance		.05		.01		NS				NS
German (College II)										
Withholding (3 weeks)	22	63	23	23	23	24	-	-	23	3.5
No Withholding	49	61	54	24	47	25	-	-	56	3.5
Significance		NS		NS		NS				NS

(1) *Mean test scores and average grades are raw scores for 1960-61, but for 1961-62 and 1962-63 have been adjusted to take account of differences in pre-test scores on SAT and MLAT.*

(2) *Numerical equivalents of grades: F=1, D=2, C=3, B=4, A=5.*

where orthography was not withheld, a large number of other factors inescapably enter into the picture. On the chance that some striking results might be obtained, however, it has seemed worthwhile to make such comparisons. In Table 3 results are shown for the first semester of 1961-62 in two languages, French and Spanish. In French the test scores of all the students at College I, where orthography was withheld for the first five periods, are compared with the scores of a class at College II where orthography was not withheld. Although the students who had access to the printed word from the beginning made test scores in reading that were higher than those of the students at College I, the difference in the averages (statistically significant at the 0.05 level) seems unimportant. There is an interesting contrast in Spanish, however, between College III and College IV. The slight difference favoring College IV, where orthography was not withheld, over College III is not significant, but the students at College III were superior to those of College IV in listening comprehension but notably inferior in reading. No other cases of significance at the

TABLE 3

AVERAGE TEST SCORES FOR TWO LANGUAGES FOR COLLEGES WITHHOLDING ORTHOGRAPHY COMPARED WITH NO WITHHOLDING

Year and Language	Speaking		Listening		Reading	
	No.	Mean	No.	Mean	No.	Mean
1961-62 (First Semester)						
FRENCH						
Withholding (College I, five periods)	114	51	119	45	120	36
No withholding (College II)	15	49	20	45	24	37
Significance		NS		NS		.05
SPANISH						
Withholding (College III, two weeks)	10	50	39	39	39	31
No withholding						
(College IV)	23	51	24	36	25	39
(College V)	21	49	22	39	21	32
(College VI)	13	46	14	40	14	33
Significance						
(III vs. IV)		NS		.01		.01
(III vs. V)		NS		NS		NS
(III vs. VI)		NS		NS		NS

0.05 level or better emerges from these comparisons.

Another type of comparison is given in Table 4 for the results of testing at year-end in 1961-62 and 1962-63. For each language shown the results for all colleges withholding are pooled and contrasted with appropriate comparison groups from one or more colleges where orthography was not withheld. In one case -- German in 1961-62 -- this resulted in a comparison of one college with another, but in most of the comparisons there is pooling of information from different colleges for both the withholding and no withholding groups. The results that are significant do not form a consistent pattern. In 1961-62, the no withholding group in French did better than the withholding group in reading, but the withholding group in German made better average scores in listening comprehension. This pattern is consistent with the idea that withholding improves the audio-lingual skills. It is found again in Spanish in 1962-63, where the withholding group was better on the speaking and listening tests, but was distinctly inferior in reading. In French in 1962-

TABLE 4

AVERAGE TEST SCORES BY LANGUAGE FOR ALL COLLEGES PARTICIPATING
ORTHOGRAPHY WITHHELD COMPARED WITH NO WITHHOLDING

Year and Language	Speaking		Listening		Reading		Writing		Grades	
2nd Semester, 1961-62	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
French										
Withholding	103	50	109	41	108	37	—	—	111	3.60
No withholding	11	50	135	41	24	41	—	—	26	3.70
Significance		NS		NS		.01	—	—		NS
German										
Withholding	—	—	18	48	18	44	—	—	18	4.00
No withholding	—	—	14	44	14	44	—	—	15	3.90
Significance	--	—		.01		NS	—	—		NS
1962-63										
French										
Withholding	33	50	34	24	39	30	13	46	40	3.48
No withholding	62	55	67	30	66	37	9	50	71	3.28
Significance		.05		.01		.01		NS		NS
German										
Withholding	21	63	22	23	22	24	—	—	22	3.50
No withholding	49	61	54	24	47	25	—	—	56	3.52
Significance		NS		NS		NS	—	—		NS
Spanish										
Withholding	41	59	42	23	43	27	42	76	43	3.58
No withholding	36	47	39	21	39	27	23	85	38	3.21
Significance		.01		.05		NS		.01		NS

63, however, this pattern did not appear. The no withholding group was superior in all of the statistically significant comparisons, speaking, listening and reading.

The statistical results of the experiments in withholding orthography are certainly not clear-cut, and it is probably too much to expect that they should be otherwise. It would be surprising if withholding periods ranging from five class meetings to nine weeks would show consistently favorable results on tests given at the end of the first semester and even more surprising if such consistent results appeared in the year-end testing program.

In addition to the statistical results there were the reactions of instructors to these experiments. Some instructors concluded that withholding is not practical at the college level and with the instructional materials that they preferred to use. Other instructors are convinced that withholding is desirable. Two of the colleges participating in this program had been withholding orthography before these experiments began and they continue to do so. In one of these colleges orthography is withheld in all beginning languages, and in the other in French only. The period of withholding is two weeks as a minimum, but in the college with the cross-the-board practice, individual teachers may go up to three weeks at their discretion.

The head of the French department in the second college was convinced that two weeks was the optimum period of withholding at the college level. He believes that some withholding is very effective in getting the student started in the sounds of the new language without the interference of the letter symbol-English sound associations habitual to him. On the other hand, he feels that the procedure cannot be maintained longer; the purely audio-lingual approach is analogical and motor in character, whereas the student by the time he enters college is firmly conditioned to an analytical and visual approach. Furthermore, the audio-visual approach needs a great deal of support which is not adequately given by presently available materials. These inhibiting conditions probably reinforced the traditional emphasis on a grammar-reading orientation (even when modified by audio-lingual techniques) in which most language teachers have been trained, and so account for the reluctance of most of the ACM colleges to adopt the withholding introductory period in beginning languages.

III The Testing Program

The Uses of Testing

The examination process by which students are called upon to demonstrate their abilities has its critics. There are undoubtedly good and bad tests and they are used sometimes well and sometimes badly, but they are an indispensable part of most programs of language instruction. They serve a number of important purposes.

In many situations tests provide valuable motivational devices. The knowledge on the part of the student that he must pass a test may lead him to intensify his efforts. The test is also a means by which the instructor conveys information to the student. Inclusion in a test dramatizes an item or aspect of the course. The test may, therefore, be a means by which the instructor indicates to the student what he considers to be significant. It is also a means by which the student is informed of the standards of performance that are expected and of his own level of achievement related to these standards.

Tests are also useful because of the information they make available to the instructor. They indicate the degree of success of instructional techniques, they reveal aspects of the course with which students generally are having difficulty, and they may indicate the specific problems encountered by individual students. They also provide a basis for judging the effectiveness of different instructional techniques or experiments.

Tests are useful, not only as used in particular courses, but in a broader setting as well. They may be a means of judging the proficiency of persons who have had a variety of previous learning experiences; they may be used for placement in college courses of persons who have already had some language instruction. They can be used to compare the outcomes of different approaches to language learning, and they can be used to determine whether individuals have achieved a desired level of proficiency such as that specified by a college foreign language requirement, or whether they are qualified for tasks that call for language competence, for research using foreign language materials, for translating, or for teaching.

The wide variety of purposes and situations in which tests are used calls for an enormous number of testing instruments and procedures. A very large number of different tests were given in the classes that participated in this program. There were certain tests, however, that were selected to serve the broad purposes of the study, and that constituted the testing program.

All the colleges that are members of the Associated Colleges of the Midwest are members of the College Entrance Examination Board and require the CEEB Scholastic Aptitude Test of applicants for admission. Scores for both the verbal and mathematical sections of the SAT were available for all students involved in the program and were used for pretesting purposes.

The Modern Language Aptitude Test was also given as a pretest. This test was developed by John Carroll and Stanley M. Sapon at Harvard University and is designed to measure a number of characteristics that are related to ability to learn a foreign language. Post tests results were adjusted on the basis of MLAT scores in the effort to eliminate the effects of differences in language learning aptitude from the results.

The process of learning a language is a sampling process. No one can know and skillfully use more than a fraction of the words, syntax, idioms and other elements of a complex language system. The sampling, however, is never random. The contrast between the reading emphasis and the audio-lingual emphasis is a contrast between sampling systems. Even within these broad categories, each instructional program involves a sampling process that reflects judgments as to the order and manner in which the various elements may be most effectively presented. Unless the student achieves complete mastery of the materials that have been presented to him, his competence will relate to only a part -- a sample -- of the whole. This competence in turn may be partial; the words and structures he can understand in listening or reading may be much more numerous than those he can use in speaking or writing.

The process of testing is again a sampling process -- but not a random one. The nature of the sampling will be determined by the purposes of the test constructor. When he is preparing a test to cover course materials he may be concerned to include those items the importance of which he wishes to impress on the student. The constructor of such a test might have reason to be gratified if every student made a perfect score. In preparing an aptitude test, he will seek items which seem to be predictive, that correlate with degree of success in learning a language. In such a test items that everyone (or no one) answers correctly are useless. A test designed to measure the relative effectiveness of instructional procedures should include only items that were included in the procedures and should include items that will discriminate effectively at both ends of the achievement scale. Tests that are designed to determine qualifications for certain types of activities -- such as teaching or translating should include items that are important in such activities. Discrimination at the extremes of performance may not be so important as discrimination within the range where the cut-off line between qualified and unqualified is likely to be drawn. Ideally such an examination would contain only items that are predictive of success in the activity.

Looked at from one point of view tests used for placement purposes are qualification tests and are used to predict how well a student may be expected to perform in the course alternatives available for placement. From another point of view such a test is diagnostic; it reveals the student's strengths and weaknesses and thus makes it possible to prescribe special activities that may facilitate his adjustment to the work of the class to which he is assigned.

Ideally in each testing situation the instrument used is designed to meet the particular need. In practice this is frequently not feasible. The construction of a reliable and valid test is a difficult and time-consuming process. It requires high competence both in the subject being tested and in the techniques of test construction. In this study it was not possible to develop specific tests for every situation. Furthermore, the collaboration of different instructors and different colleges meant that some uniformity in the testing program was necessary. Such uniformity, however, necessarily meant compromises. In the experimental program, for example, it was not feasible to standardize all of the materials and methods in the class sections participating in the experiment, and it was not possible to develop examinations that sampled only such a common background. It was important to make use of tests that had been scientifically developed and extensively pre-tested even though the subject matter coverage they presumed did not exactly coincide with the coverage at any of the colleges.

The Tests Used

When the program was started it was hoped that the new language tests being developed in the Cooperative Test Division of the Educational Testing Service in cooperation with the Modern Language Association of America and the United States Office of Education would soon be available. These tests were to measure all forms of the basic language skills and their development had been under way for 5 years and had involved the collaboration of 75 language and test experts. These tests were not published for general use until 1964, but it was possible to participate in the pre-testing program for them in the second semester of the third year of the program, 1962-63.

In the first two years the tests that were used for post-testing were the College Entrance Examination Board's Achievement Tests from earlier years that had been made available for use by educational institutions generally. They were available in French, German and Spanish. There are two sections. The first requires one hour for administration and emphasizes reading and comprehension and understanding of language structure. The second part is a thirty-minute listening comprehension test. In terms of the four basic language skills, these two parts may be conceived of as sampling reading and listening skills respectively.

Because of the audio-lingual emphasis of much of the activity in this program it was highly desirable to start testing the speaking skill as soon as possible. The new Cooperative tests mentioned above were to include a speaking test, but were not to become available until sometime in the academic year 1962-63. It was therefore, decided to construct speaking tests in four languages at the Faculty Summer Workshop in 1961 to be used in 1961-62. Special committees were set up for each language and the tests prepared were reviewed and discussed by other faculty members present at the Workshop.

The "French Speaking Proficiency Test,"[1] prepared by Paul Pimsleur was used as a model and Professor Pimsleur attended the workshop and served as a consultant on speaking test construction. Two tests were prepared: one for administration at the end of the first semester of the elementary course and the other for administration at the end of the second semester. A careful study was made of the teaching materials to be used at the member colleges during 1961-62 and the test items were directly related to these materials. This met in part the sampling problem discussed above and provided tests appropriate to the measurement of the results of the experiments to be conducted during the year. The process of administering and scoring the tests at the end of the first semester suggested a number of changes. The proposed second semester tests were, therefore, thoroughly revised before they were administered. Copies of all the tests and of the scoring sheets used may be found in Appendix 1.

---

[1] See Paul Pimsleur, "A French Speaking Proficiency Test", The French Review, Vol. XXXIV, No. 5, (April 1961) pp. 470-479.

The tape recorder provides a facility by which it is possible to administer a speaking test under reasonably uniform conditions. A laboratory in which each student's responses can be separately recorded makes it possible to obtain a tape for each student which can then be graded in accordance with pre-determined standards. Unfortunately, at the end of the first semester of 1961-62 several of the colleges did not have laboratories of the requisite flexibility and, at those colleges it was necessary for faculty members to administer the examination and score the students as they responded.

It is possible to grade a written test in much less time than was devoted to its composition by the student and the contrast between the time spent on the examination by the student and the time spent by the grader is even more drastically reduced for objective tests. Indeed, machine scoring can reduce the grading of objective tests to a purely mechanical operation. With a speaking test, on the other hand, the grader must listen to everything said by the student and the time spent in grading the test will, therefore, be determined by the time each student devotes to the test multiplied by the number of students being tested. It is consequently important to have a test that is no longer than necessary to meet the intended purposes. Fortunately, it is possible in a very few minutes to present a student with speaking tasks that will reveal a great deal with respect to his ability to speak a foreign language. In a carefully designed procedure for administering the ACM French Speaking Proficiency Test "live", Professor Loring Knecht of St. Olaf College found that it took just under eight minutes to administer the First Semester French test to one student. The Second Semester Tests were put on tapes exclusively and each required 20 minutes.

Speaking tests were prepared for Russian and administered at the end of the first semester. The number of students taking this test, however, was so small that it did not seem worthwhile to revise it or, indeed to administer it after the first trial. No further information, therefore, is presented with respect to it.

At the end of the first semester in 1961-62 the specially designed speaking tests were administered at eight colleges.

Because of limitations in laboratory space and equipment, it was necessary for instructors to administer most of these tests to students individually, "live". Table 5 shows the number of tests administered in this manner at each college and the number administered by use of a tape recorder. Altogether these tests were administered to 1,223 students, 700 of whom took them"live."

The Second Semester Speaking Tests were administered by tape only. They were voiced by native speakers and the student responses for each language were all graded by a single native speaker.

In 1962-63 the new Cooperative Tests became available and were used in the test administrations at the end of that year. These tests measure the four basic language skills, reading, writing, listening and speaking.

The Results of Testing: Elementary Level

The detailed results of the pre-testing and post-testing programs are presented by college, by language and by sex in Appendix 2. The number of students taking the

TABLE 5

METHOD OF ADMINISTERING ACM SPEAKING TESTS in January, 1962

(Numbers of Students)

COLLEGE	FRENCH		GERMAN		SPANISH	
	Live	Tape	Live	Tape	Live	Tape
Beloit	63	—	—	—	25	—
Coe	20	40	24	—	21	90
Cornell	—	59	—	30	25	—
Grinnell	—	—	18	—	—	—
Knox	—	42	—	58	9	—
Lawrence	20	50	19	90	49	—
Monmouth	56	—	85	—	78	—
St. Olaf	75	64	76	—	37	—
TOTALS	234	255	222	178	244	90

tests and their mean scores on the different tests are shown in Table 6.

Because different tests were given in the two years, it is not possible to make comparisons of the results from one year to the next. The only valid comparison provided by Table 6 is the contrast between the male and female students. The men outperformed the women on the Scholastic Aptitude Test in mathematics, but the women equalled or outscored the men on all the other tests given.

The data presented in Appendix 2 supply another basis for comparing the performance of males and females. It is possible to compare the scores of men and women on each test in each language at each college. When this is done and the comparisons tabulated in terms of the number of instances in which one sex scored higher than the other, we obtain the results shown in Table 7. It is immediately clear that in matters relating to language the women outperform the men. In mathematics as measured by the Mathematics SAT, the men have an advantage, outscoring the women in 35 instances to 2. In other comparisons, all of which

TABLE 6

General Summary of Test Results

At Elementary Level 1961-62 and 1962-63

	1961-62				1962-63			
	Male		Female		Male		Female	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean
PRE-TESTS								
All Participants								
SAT Math	551	564	486	508	438	582	363	514
SAT Verbal	551	528	485	540	438	539	363	547
MLAT	562	115	517	126	281	113	261	129
POST-TESTS								
French								
Speaking	170	489	239	492	127	50	146	51
Listening	229	451	284	456	138	27	160	29
Reading	237	363	287	368	137	32	161	34
Writing	-	-	-	-	91	56	125	67
Course Grade	248	3.5	291	3.8	147	3.1	167	3.7
German								
Speaking	102	492	74	503	185	56	115	59
Listening	171	427	119	435	207	26	131	28
Reading	168	393	121	395	200	27	129	29
Writing	-	-	-	-	162	71	49	80
Course Grade	182	3.5	123	3.8	215	3.3	135	3.9
Spanish								
Speaking	122	479	70	499	107	51	104	55
Listening	178	397	122	410	114	19	108	23
Reading	175	327	120	341	108	23	100	27
Writing	-	-	-	-	100	69	92	76
Course Grade	194	3.2	125	3.8	121	3.1	108	3.5

TABLE 7

Relative Standing of Males and Females on Pre-tests and Post-tests for Groups Consisting of Those Tested at Each College in Each Language 1961-62 and 1962-63

Test	Number of groups in which the mean score was:		
	Higher for Males	Higher for Females	Tied
SAT Math	35	2	1
SAT Verbal	12	26	
MLAT	3	31	
Speaking	5	25	3
Listening	7	30	1
Reading	3	34	
Writing	3	13	
Course Grades	3	34	1

are related to language facility, the advantage is with the women. This is least pronounced in the SAT verbal test, where the score is 26 to 12. Next in order comes the listening comprehension test where the score was 30 to 7, with one group tied. In the MLAT, the other achievement tests, and course grades the superiority of the women is quite marked.

Additional summary information of interest is provided by the correlation coefficients between scores on the various tests. These correlations are presented for the first semester of 1961-62 in a correlation matrix for each language.

The first matrix to be considered related to the performance of students in elementary French courses in the first semester of 1961-62. It is shown in Table 8. It reveals significant correlation between the pre-tests: 0.43 between the two SAT tests, 0.40 between the mathematics SAT and the MLAT and 0.49 between the verbal SAT and the MLAT. To some extent the pre-tests were measuring the same thing. This was presumably "aptitude" for academic work.

TABLE 8

Correlation Matrix for French Students

End of First Semester, 1961-62

	SAT Math.	SAT Verbal	MLAT	Speaking	Listening	Reading
SAT Verbal	.43					
MLAT	.40	.49				
Speaking	.05	.18	.28			
Listening	.05	.14	.12	.18		
Reading	.10	.22	.28	.29	.20	
Course Grades	.16	.23	.42	.55	.20	.38

N =400

They were, however, very poor predictors of academic success as this was revealed by the scores made on the post-tests. It is perhaps not surprising that the Mathematics SAT did not predict success in language learning very well, but the Verbal SAT was only slightly better. The MLAT was somewhat more effective than the Verbal SAT particularly as a predictor of course grades, for which the coefficient of correlation was 0.42. The most striking coefficient in the correlation matrix was the 0.55 obtained between the score on the ACM Speaking Test and course grades.

The situation revealed by the correlation matrix for German students (Table 9) is similar to that for French students in that there is a moderate correlation between the pretests, negligible or very low correlations between the pretests and post-tests, and an appreciable correlation between the ACM Speaking Test and course grades (a coefficient of 0.53). It differs, however, in the higher intercorrelations between the post-tests. These were:

Between:

Speaking and listening comprehension	0.51
Speaking and reading	0.50

TABLE 9

Correlation Matrix for German Students

End of First Semester, 1961-62

	SAT Math.	SAT Verbal	MLAT	Speaking	Listening	Reading
SAT Verbal	.46					
MLAT	.31	.45				
Speaking	-.01	.15	.28			
Listening	-.01	.10	.16	.51		
Reading	.13	.24	.24	.50	.45	
Course Grades	.13	.10	.34	.53	.38	.43

N = 440

Speaking and course grades	0.53
Listening comprehension and reading	0.45
Listening comprehension and course grades	0.38
Reading and course grades	0.43

The correlation matrix for Spanish students is given in Table 10. It resembles that for German students in its main features. The coefficient of correlation between the ACM Speaking Test and course grades (0.53) is in this case slightly surpassed by the correlative between listening comprehension and reading and between reading and course grades. Both of these were 0.58.

Two participants in this program, Klaus A. Mueller, the Coordinator of Language Instruction Programs, and William Wiersma, Consultant in Psychometrics, reached the following conclusion with respect to the relatively high correlation between the ACM Speaking Test and the course grades at the end of the first semester of 1961-62:[1]

---

[1] "Correlation of Foreign Language Speaking Competency and Grades in Ten Midwestern Liberal Arts Colleges," The Modern Language Journal, Vol. XLVII, No. 8, Dec., 1963, pp. 353-355

TABLE 10

Correlation Matrix for Spanish Students

End of First Semester, 1961-62

	SAT Math.	SAT Verbal	MLAT	Speaking	Listening	Reading
SAT Verbal	.43					
MLAT	.38	.48				
Speaking	-.03	.06	.19			
Listening	.14	.33	.31	.38		
Reading	.13	.29	.35	.41	.58	
Course Grades	.05	.12	.39	.53	.38	.58

N = 350 For the speaking test N = 230

> "This suggests that speaking ability is emphasized in foreign language instruction in our colleges and relative competence in speaking is reflected in the course grades assigned by instructors. It further suggests that proficiency in audio-lingual skills as well as in reading and writing is an important objective of language instruction in these colleges and that students' proficiency in speaking the foreign language may have greater weight than any other factor in determining success or failure in the language course."

An alternative hypothesis may be advanced in the light of the discussion earlier in this chapter concerning the sampling nature of the testing process. The Speaking Test and the course grades are the only measures that were specifically related to the samples of the languages represented by the learning experiences of the students up to that point. Since those tests and grades were given after only one semester of study at the elementary level, the students' experiences were relatively small samples of the total language. In other words these two post-experience measures were the only ones directly related to the particular experiences these students had had with these foreign languages. It is not surprising, therefore, that the correlation between them is higher for the French and German students than other correlations. It may be that the higher intercorrelations between post-tests for the Spanish students reflects a greater congruence between the samples represented by various post-tests and course grades.

It is possible to test this alternative hypothesis. It was advanced before correlation matrices for 1962-63 had been prepared. The testing program in the later year differed significantly from that for 1961-62. The new MLA Cooperative Language Tests were used, including tests of writing and speaking. The tests were administered at the end of a year of college study -- rather than at the end of one semester. The course grades were, therefore, the only measure that related directly to the sample of the language represented by the students' learning experiences, and the sample of the language tested should be twice as large as in the earlier year. One should expect for the second year, that the correlation matrix would more closely resemble that for Spanish students than that for French or German students. There should be fairly good correlations between the post-tests and no one of them should correlate in an exceptional manner with course grades.

The correlation matrices for 1962-63 presented in Tables 11, 12 and 13 may be examined with reference to the hypothesis presented. The correlation matrix for French students (Table 11) shows relationships among the pretests very similar to the French matrix for 1961-62 (Table 8). The pre test intercorrelations were: 0.43, 0.40 and 0.49 in 1961-62 and 0.41, 0.32 and 0.46 in 1962-63. The intercorrelations between the post-tests are definitely higher in the second year. This may best be seen by setting these correlation coefficients for the two years side by side. The correlation between speaking and course grades is the lowest shown in the second year, although the differences in

TABLE 11

Correlation Matrix for French Students

End of 1962-63

	MATH		VERBAL		MLAT		SPEAKING		LISTENING		READING		WRITING	
	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio
VERBAL	269	.41												
MLAT	153	.32	153	.46										
SPEAKING	234	-.01	234	.20	140	-.08								
LISTENING	254	.17	254	.40	148	.10	270	.60						
READING	252	.25	252	.41	145	.39	262	.37	289	.58				
WRITING	208	.18	208	.35	149	.57	200	.39	217	.61	212	.66		
GRADE	268	.20	268	.21	155	.25	274	.36	299	.40	299	.47	221	.50

the coefficients for speaking with reading, writing and course grades are minor. The relatively high correlation between speaking and listening seems to suggest a relationship between the audio-lingual skills, but this relatively high relationship is not found in the correlations for German and Spanish students. In the second year a new post-test was introduced -- writing. It was modestly correlated with speaking, but the highest correlations shown were with this post-test and listening and reading.

	Speaking		*Listening*		*Reading*		*Writing*
	1961-62	*1962-63*	*1961-62*	*1962-63*	*1961-62*	*1962-63*	*1962-63*
Listening	*.18*	*.60*					
Reading	*.29*	*.37*	*.20.*	*.58*			
Writing	-	*.39*	-	*.61*	-	*.66*	
Course Grade	*.55*	*.36*	*.20*	*.48*	*.38*	*.47*	*.50*

The matrix for German Students in 1962-63 (Table 12) shows the usual modest correlations between the pre-tests. The intercorrelations among post-tests are compared for the

TABLE 12

Correlation Matrix for German Students

End of 1962-63

	MATH		VERBAL		MLAT		SPEAKING		LISTENING		READING		WRITING	
	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio
VERBAL	310	.34												
MLAT	235	.21	235	.36										
SPEAKING	260	.04	260	.11	199	.23								
LISTENING	295	.03	295	.20	229	.22	297	.48						
READING	285	.10	285	.26	227	.41	285	.40	322	.62				
WRITING	239	.22	239	.21	228	.32	225	.61	256	.52	257	.67		
GRADE	305	.09	305	.15	238	.33	298	.46	336	.52	326	.50	259	.54

two years below:

	Speaking		*Listening*		*Reading*		*Writing*
	1961-62	*1962-63*	*1961-62*	*1962-63*	*1961-62*	*1962-63*	*1962-1963*
Listening	.51	.48					
Reading	.50	.40	.45	.62			
Writing	-	.61	-	.52	-	.67	
Course Grade	.53	.46	.38	.52	.43	.50	.54

The contrast between the two years is not as pronounced as for the French students. All the correlations with speaking are lower in the second year, but the other post-test correlations are higher. There is no evidence that the four basic language skills divide into an audio-lingual group and a "literate" group. The correlations of writing with speaking (0.61) and listening (0.52) are higher than the correlation between the two audio-lingual skills (0.48).

	Speaking		*Listening*		*Reading*		*Writing*
	1961-62	*1962-63*	*1961-62*	*1962-63*	*1961-62*	*1962-63*	*1962-1963*
Listening	.38	.32					
Reading	.41	.29	.58	.46			
Writing	-	.42	-	.41	-	.67	
Course Grade	.53	.37	.38	.33	.58	.41	.51

For Spanish students (Table 13) all the coefficients between post-tests are lower for 1962-63 than they were for 1961-62. This is in contrast to the correlations for French students. It is possible to characterize the contrasting situations between the languages by pointing out that of the six post-test comparisons: in French four coefficients were higher in the second year, two were lower; in German three were higher, three were lower; in Spanish all six were lower.

When we look at these results in terms of the hypothesis that the relatively high correlations between speaking and course grades in 1961-62 were due to the greater relevance of these two measures to the students' language experience in that year, we find support for the hypothesis. In the second year there were ten intercorrelations between post-tests. For French students all nine of the other correlations exceeded that for speaking and course grades. In German eight of the other correlations were higher and in Spanish, six.

TABLE 13

Correlation Matrix for Spanish Students

End of 1962-63

	MATH		VERBAL		MLAT		SPEAKING		LISTENING		READING		WRITING	
	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio
VERBAL	220	.47												
MLAT	143	.21	143	.24										
SPEAKING	201	.06	201	.21	134	.18								
LISTENING	212	.10	212	.27	141	.27	207	.32						
READING	198	.06	198	.29	125	.48	193	.29	203	.46				
WRITING	185	.11	185	.18	125	.30	178	.42	187	.41	190	.67		
GRADE	218	.16	218	.13	144	.23	209	.37	219	.33	206	.41	190	.51

The subsidiary hypothesis that the students' longer experience with the language in the second year should lead to higher intercorrelations between post-tests is not confirmed. The results for French students conform to this pattern, but those for German and Spanish students do not.

The Results of Testing: Advanced Level

A most important battery of tests is the Modern Language Association Language Proficiency Tests for Teachers and Advanced Students. These are designed to measure language skills at a high level as well as knowledge of linguistics, of the civilization and culture of the countries where a given foreign language is spoken and of professional preparation for teaching the language. These tests were under development when the program began. Preliminary forms of the tests were administered in 1959 at Summer Language Institutes for secondary school teachers conducted by colleges and universities under contracts with the United States Office of Education. They were also taken by faculty members who attended the ACM Faculty Workshops in the summers of 1960, 1961, and 1962.

The skills section and the Civilization and Culture section were taken by a total of 27 students at four ACM colleges in the spring of 1962. In May of 1963 all the ACM colleges participated in the administration of the final form of these tests to language majors. In all 173 students took these tests in 1963, most of whom were seniors; all of them had completed at least three years of language study on the college level. Relatively few students took the linguistics and professional preparation tests.

The results of this advanced-level testing are presented in Table 14. For comparative purposes the mean scores made in 1963 by language majors at colleges and universities in Pennsylvania and by secondary school foreign language teachers at Summer Language Institutes are included.

The results in French are shown graphically in Chart 1. For each test the mean scores of five different groups are shown. The first bar on the left represents the mean score of the ACM students taking the test in 1962. The second bar gives the mean score for ACM students taking the test in 1963 and the third bar represents the performance of

TABLE 14

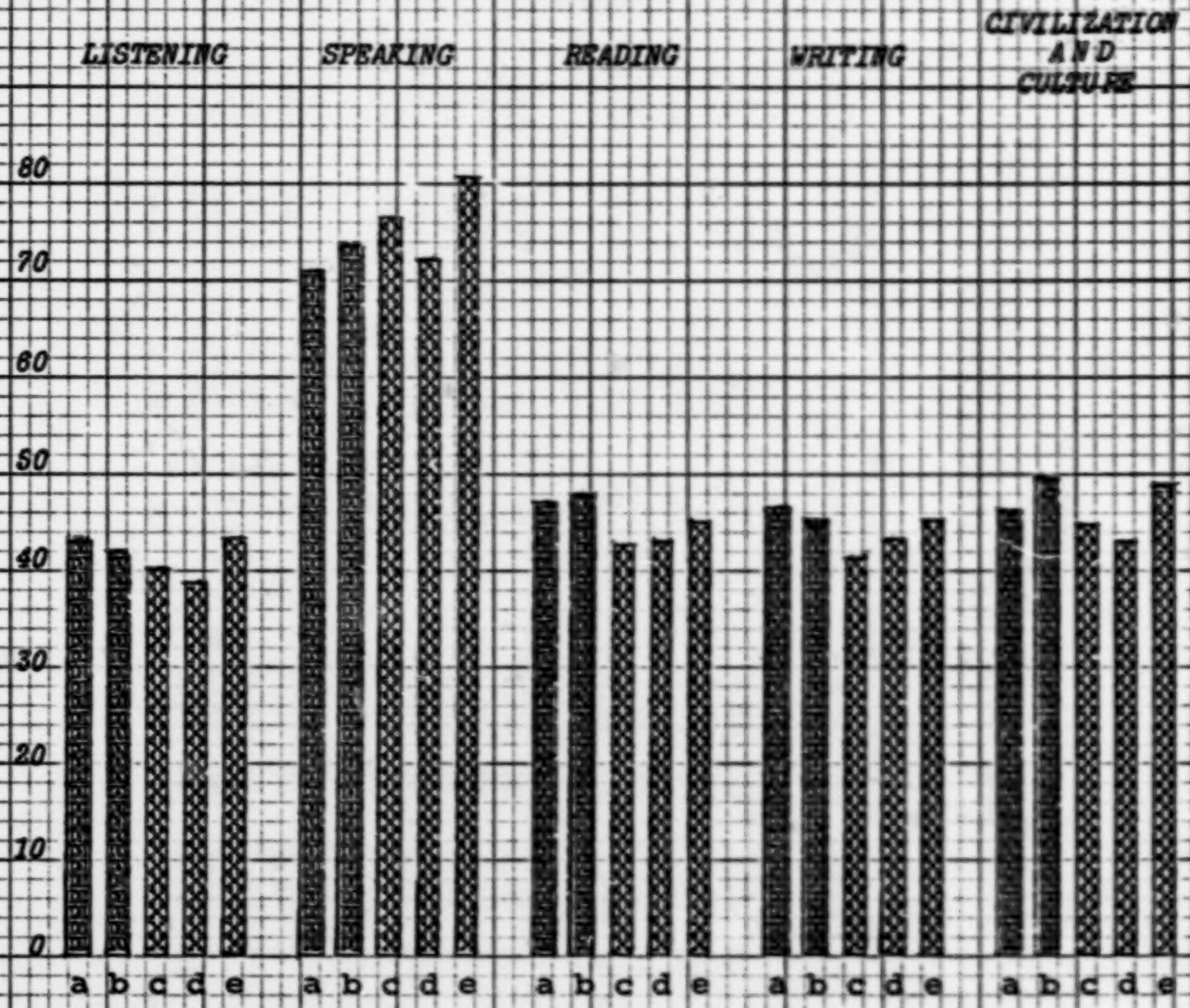
MEAN SCORES ON MODERN LANGUAGE ASSOCIATION LANGUAGE PROFICIENCY TESTS

	ACM Students									
	1962		1963		Pennsylvania Students		Teachers Pre-test		Teachers Post-test	
TEST	No.	Mean	No.	Mean	Number	Mean	Number	Mean	Number	Mean
French										
Listening	13	42.4	72	41.7	162	38.7	4720	38.1	4770	42.8
Speaking	13	70.9	66	73.2	157	76.3	4494	72.1	4577	81.0
Reading	13	46.6	71	47.8	162	42.9	4722	43.3	4770	45.7
Writing	13	46.2	72	45.1	161	41.3	4723	43.0	4769	45.7
Civ. & Culture	13	46.9	52	48.7	47	44.6	4640	43.6	4611	49.3
German										
Listening	6	40.2	38	40.3	55	38.9	1102	38.4	1256	42.1
Speaking	6	74.0	37	81.7	52	82.2	1086	79.5	1232	85.3
Reading	6	48.2	39	49.4	55	44.8	1103	44.4	1256	48.1
Writing	6	53.3	33	52.6	55	45.0	1103	45.8	1258	48.3
Civ. & Culture	6	53.3	33	52.0	28	48.7	1101	46.7	1178	53.4
Spanish										
Listening	6	45.2	48	40.7	136	39.1	4394	39.3	4400	41.8
Speaking	5	96.2	48	64.0	130	76.2	4271	72.1	4316	78.2
Reading	6	51.7	48.	44.3	138	41.6	4446	42.2	4344	44.6
Writing	6	56.3	47	44.2	137	45.3	4444	46.2	4547	50.2
Civ. & Culture	6	55.2	47	52.7	36	50.5	4342	49.6	4443	56.2

CHART 1

MEAN SCORES ON MODERN LANGUAGE ASSOCIATION LANGUAGE PROFICIENCY TESTS

FRENCH



a. ACM Students, 1962
b. ACM Students, 1963
c. Pennsylvania Students
d. Teachers Pre-test
e. Teachers Post-test

the Pennsylvania language majors in the same year. The fourth bar shows the mean score achieved by participants in the Language Institutes on pre tests and reflects their performance at the beginning of their Summer Institute experiences. The final bar on the right shows the mean score achieved by Institute participants at the end of the Institute.

Except for the speaking test, there is a general similarity in the patterns of performance shown by the different groups. This pattern is characterized by a roughly U-shaped design formed by the tops of the bars. The two ACM student bars on the left and the right-hand bar for the Institute post-tests tend to rise higher than the bars for the Pennsylvania students and the Institute pre tests.

It is clear that the Institute program resulted in marked gains in all the language skills. It is also clear that in all the skills tested except speaking the ACM students make relatively good showings. The results on the speaking test will be discussed after data have been presented on the results of these tests in other languages.

The results for German are presented in Chart 2. It is immediately apparent that they are similar to the results for French. The U-shaped conformation is evident for all the tests except for Speaking. In some cases, as for Listening, the U is rather flat and the Institute post-test averages are higher than for either year of the ACM student testing. In other cases, as in Writing, the contrasts are more pronounced and both of the ACM groups are superior to the Institute Post-tests. Again Speaking is an exception.

The comparison for Spanish shown in Chart 3 exhibits U-shaped conformations, but there are some striking contrasts with the two previous charts. The ACM students who took these tests in 1962 clearly outperformed all the other groups on every test except Civilization and Culture. The 1963 ACM group performed well in Listening, Reading and Civilization and Culture, but made lower scores than any other group in Speaking and Writing. It should be pointed out that the high performances recorded by the ACM 1962

CHART 2

MEAN SCORES ON MODERN LANGUAGE ASSOCIATION
LANGUAGE PROFICIENCY TESTS

GERMAN

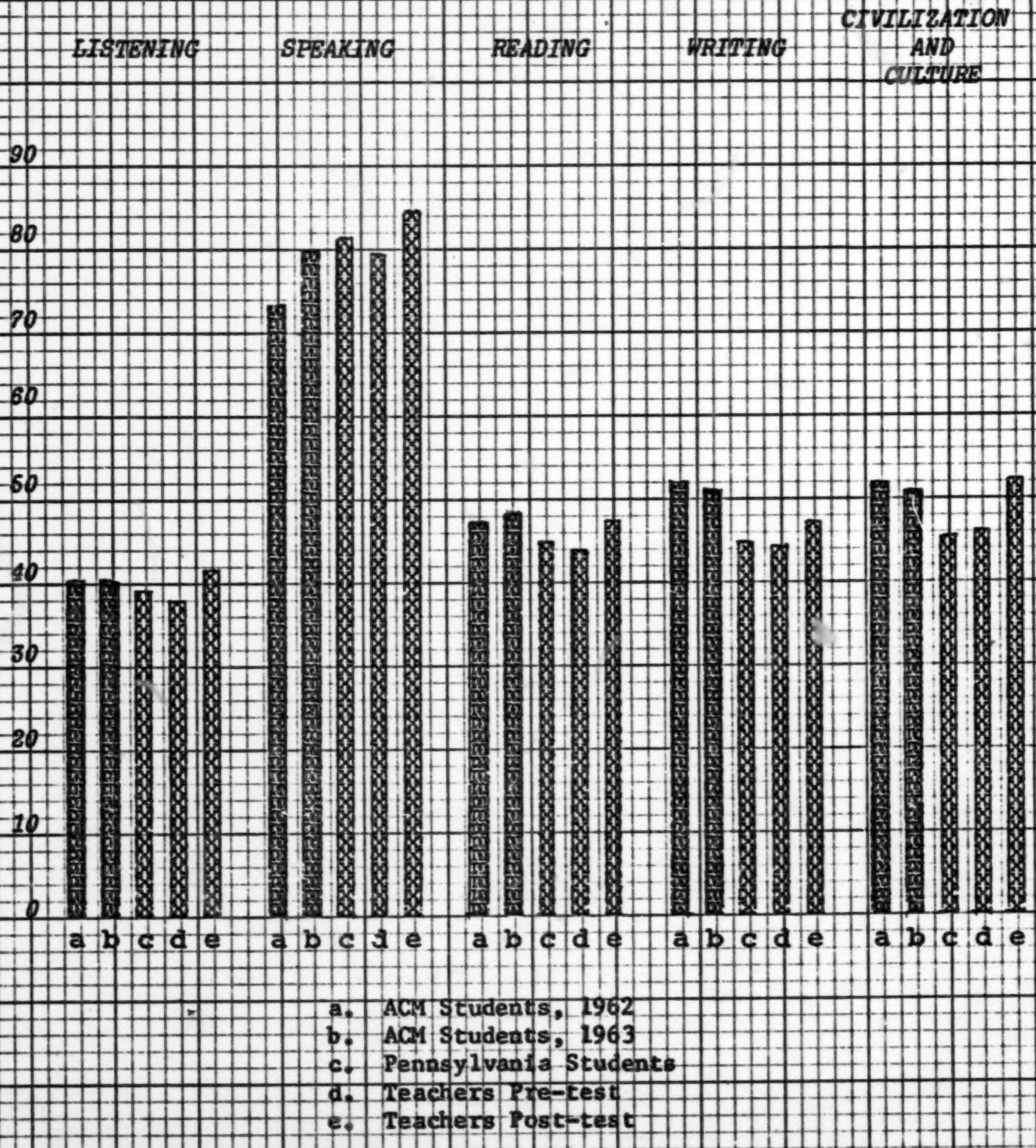
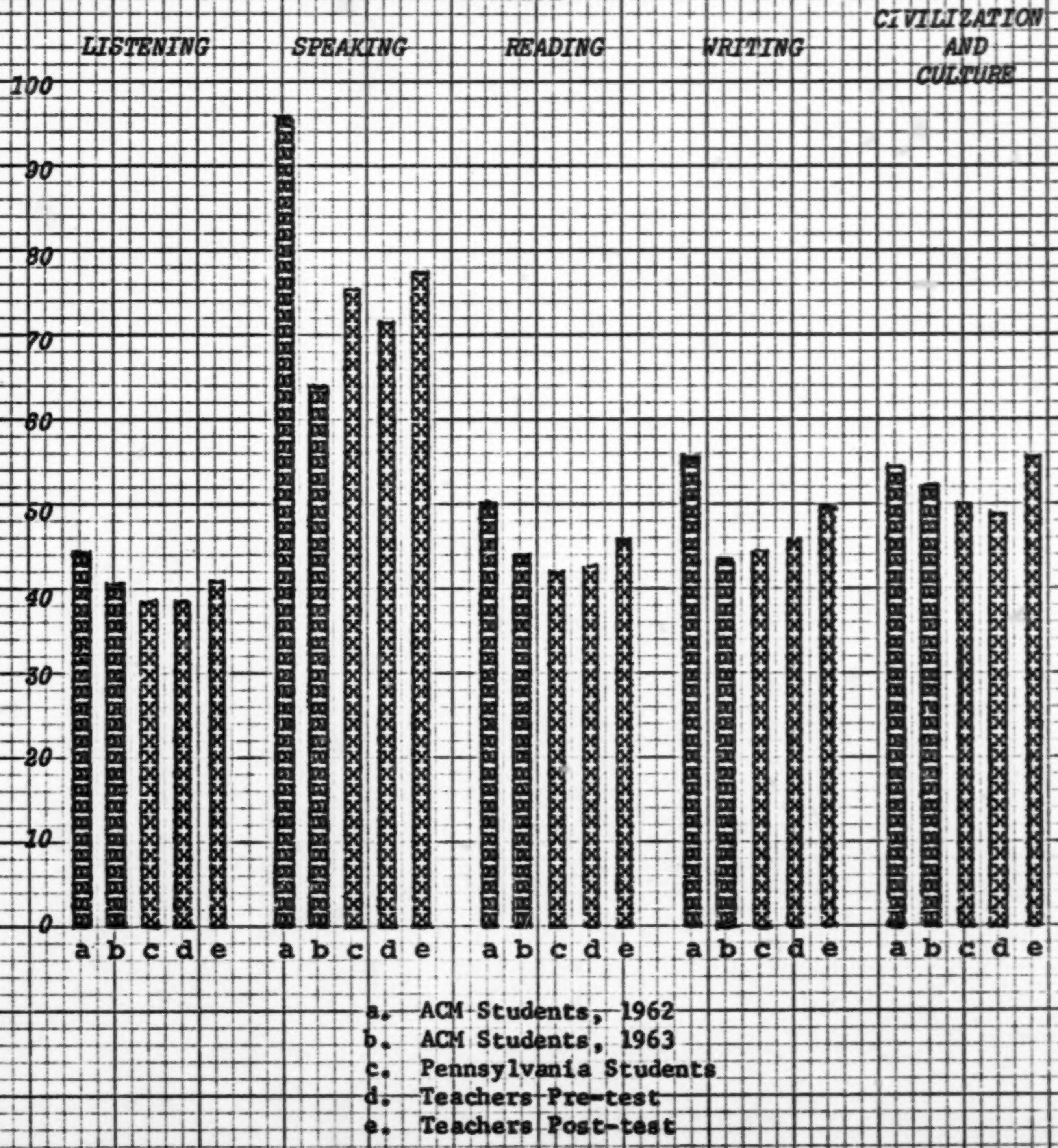


CHART 3

MEAN SCORES ON MODERN LANGUAGE ASSOCIATION LANGUAGE PROFICIENCY TESTS

SPANISH



group were averages of only five students in Speaking and six in the other tests. Of these six students, two were from Coe College, three from Monmouth and one from St. Olaf. They seem not to have been specially selected.

The test results with respect to the audio-lingual skills are anomalous. The ACM students did quite well in listening comprehension, but -- with the exception of the Spanish students in 1962 -- not so well in Speaking. It is not possible, at this time, to determine the causes of this contrast. One possible explanation is that familiarity with tape recorders and experience in recording one's foreign language utterances are more important in the speaking test than in the listening test. The ACM senior students who took these tests in 1963 were freshmen in 1959-60 when there were only two language laboratories in ACM colleges. The number of laboratories increased to 6 in 1960-61, to 9 in 1961-62; to 10 in 1962-63. In the early stages the laboratories were used largely for elementary and intermediate language instruction. It seems likely, therefore, that most of the students taking the test in 1963 had not had much experience with tape recorders.

IV, THE LANGUAGE LABORATORY

Recent Developments

Many factors have undoubtedly contributed to the language laboratory movement that has swept this country -- and other parts of the world -- in recent years. Two that may be mentioned here are the growing tendency to look at language as a system of orally produced sounds and the improvement in the fidelity of sound reproduction equipment. If language is a system of sounds it is important for the learner to have as much contact as possible with these sounds. The overworked language teacher could not possibly supply appropriate sounds in the amount neccessary to provide college language students with adequate experience with them. With equipment of good fidelity and an adequate supply of good recordings, it is possible to increase greatly the opportunities of the student for contact with the foreign idiom. It was inevitable that the production of machines and laboratories would run ahead of the development of effective instructional materials to use with the machines and the discovery of the most effective ways to use sound reproducing machinery in learning foreign languages. A major concern of this program was to attempt to appraise the usefulness of different types of laboratories, to consider different ways in which laboratories could be used and to be concerned with teaching materials that might contribute to language learning.

Equipment at ACM Colleges

A great variety of language laboratory equipment with considerable range in costs is now being offered to schools and colleges. The installations in the ten ACM colleges ranged in cost from about $170 per student position to over $750. The more expensive laboratories are described as "completely flexible." This means that at each student position it is possible to listen and to record either individually or in concert with other students. The individual student can also play back his own tape and listen to his previous utterances. The minimum cost laboratory is one in which all the students can listen through earphones to a common source and there are no microphones either for transmitting or recording the sounds they utter. As the laboratory becomes more elaborate more sound sources may be installed and microphones added to provide an "audio-active" feature in which the student speaks into a microphone and his utterances are transmitted to him through the earphones, but are not recorded. In both "flexible" and "inflexible" laboratories it is possible to have a central console at which a monitor can hear each student's speech and can communicate with the students either singly or in groups. Most of the less flexible laboratories will have a few tape recorders which permit a limited number of students to record and play back what they have spoken in the foreign language.

When the present project was initiated, only two of the member colleges, Coe and Knox, had language laboratories, and these were both small. During the course of the program all of the colleges acquired such laboratories. The size of these installations, the date of acquisition and the nature of the principal equipment are set forth in Table 15. All of the laboratories have central consoles which permit a monitor to listen to and communicate with each student or with any desired group of students.

TABLE 15

LANGUAGE LABORATORY INSTALLATIONS AT MEMBER COLLEGES

COLLEGE	NUMBER OF PLACES	DATE OF INSTALLATION	PRINCIPAL EQUIPMENT
Beloit	25	1960	Dual track tape recorders
Carleton	35	1963	Dual track tape recorders
Coe	24	1959*	Central 16-track recorders
Cornell	30	1960	Dual track tape recorder
Grinnell	24	1961	Dual track tape recorders
Knox	26	1959*	Dual track tape recorders
Lawrence	30	1961	Non-recording but audio-active
Monmouth	24	1961	Non-recording but audio-active
Ripon	35	1962	Non-recording but audio-active
St. Olaf	58	1962	Dual track tape recorders and non-recording audio-active

*Date of expansion of laboratory to present size

The installation of these laboratories reflected the conviction on the part of language faculties that such equipment could make a contribution to language learning. A major concern of this program has been to learn more about the contribution and effectiveness of different types of equipment and to discover effective laboratory procedures. An important part of the activities of the participants in the Summer Faculty Workshops concerned these questions.

As may be seen from Table 15 laboratory facilities were installed by the colleges at various times between 1959 and 1962. At the beginning some of the colleges that did not have language laboratories were interested in participating as controls in experiments on laboratory uses. As laboratories were acquired interests quite naturally shifted. As a consequence, the experimental program underwent changes. An initial experiment contrasting instruction with laboratory supplement with instruction without laboratory was dropped after 1961-62 and an experiment contrasting four types of laboratory changed to a contrast of two types. The reason for this latter change is that the actual laboratory installations quickly focused on two main types, (1) a fully "flexible" laboratory, with individual tape recorders at each student position, and (2) a non-recording laboratory where students could listen and speak into audio-active microphones, but where student recording is possible only by use of supplementary equipment -- some of which is found in every laboratory.

Table 15 shows the nature of the laboratories that eventually developed. Five of the colleges, Beloit, Carleton, Cornell, Grinnell and Knox have dual-track tape recorders at each student position. St. Olaf has two separate laboratories, one with dual-track tape recorders, the other with non-recording equipment. St. Olaf also has three completely enclosed booths that can be set up for almost any type of listening and recording activity. Coe has a centralized tape recorder capable of recording simultaneously the sounds originating in any 16 of the 24 booths. Lawrence, Monmouth, and Ripon have non-recording laboratories. This variety provided the basis for the experiment with respect to types of laboratory equipment.

Laboratory versus No-Laboratory Experiment

In the first flush of excitement over the acquisition of laboratories at some colleges and while other participating colleges did not have them, there was interest in an attempt to compare the performance of students who had access to the laboratory with the performance of those who did not. An effort was made to use similar methods and procedures (except for access to the laboratory) for the two groups. It was recognized that important variables such as college and teacher, could be controlled in only a few situations, but such an experiment could utilize even the personnel at colleges that did not yet have laboratories and was a way of insuring that everyone could be involved. It provided the basis for the accumulation of experience and evidence, and it contained the possibility of a dramatic demonstration if the hopes of the most ardent supporters of laboratories were realized.

Stated in the null form the hypothesis to be tested was: "The use of a language laboratory makes no discernible difference in the proficiencies achieved by language students in the four basic foreign language skills." Although there was some recognition of the infrequency of significant results in experiments of this type, most of the participants expected a clear disproof of the hypothesis.

The most fully controlled experiment was carried out in French at St. Olaf College in 1961-62. St. Olaf had not yet acquired the excellent and varied laboratory facilities that it now has. Some thirty students could listen to a single program source through audio-passive earphones with individual controls and respond in unison. One experimental section had four, thirty-minute, required laboratory sessions each week using this equipment. Another experimental section was supplied with portable record players which they could take to their rooms, with discs specially pressed with exercises that were used in the laboratory. No attempt was made to control the amount of time spent with the discs, but students were urged to devote as much time as possible to their use. There were three control sections. One instructor taught the laboratory section and a control section; another taught the take-home disc section and a control section and a third taught a control section only. The textbook was designed for audio-lingual instruction and was accompanied by a workbook which functioned in a fashion similar to programmed learning materials.

The results of the experiment at St. Olaf are summarized in Table 16. The ACM Speaking Test was given by individual administration at the end of the first semester. Unfortunately, St. Olaf did not have the tape recording facilities with which to administer the speaking test at the end of the second semester. The College Entrance Examination Board tests administered at the end of the second semester were higher level tests than those used the first semester. The lower scores do not represent retrogression!

Table 16 records no special effects from the use either of the laboratory or the take-home discs. There are no statistically significant differences but the differences in the means of the speaking tests suggest that Instructor A put somewhat more emphasis than the other

TABLE 16

RESULTS OF EXPERIMENTATION
with
LABORATORY and NON-LABORATORY SECTIONS
of Elementary French at
ST. OLAF COLLEGE, 1961-62

Section	Instructor	Laboratory Support	Speaking		Listening		Reading	
			No.	Mean	No.	Mean	No.	Mean
1	A							
	1st Semester	(Non-recording)	29	53	32	45	32	35
	2nd Semester	(audio passive)			30	40	30	36
2	A							
	1st Semester	None	28	53	29	44	29	35
	2nd Semester				25	39	24	35
3	B							
	1st Semester	None	31	49	32	45	32	37
	2nd Semester				29	42	29	40
4	C							
	1st Semester	Take home discs	27	48	28	46	27	36
	2nd Semester				23	38	25	37
5	C							
	1st Semester	None	27	49	24	45	25	36
	2nd Semester				22	42	21	38

instructors on this skill.

The ten colleges were obtaining various types of laboratories, and several obtained relatively inexpensive equipment in the initial phases of acquiring laboratory facilities. This suggested the possibility of comparing the results obtained from the use of different types of equipment. Such a comparison is in no sense a controlled experiment. In addition to the variations in laboratory equipment, there are also differences in instructors, colleges, teaching materials, number of class meetings and time spent in the laboratory. These variables, which are relevant to teaching language but uncontrolled in this experiment, made a neat, laboratory-type test of the null hypothesis impossible.

During the academic year 1961-62, the number of students participating in the experiment designed to appraise the impact of language laboratory instruction was about 550 in the first semester and about 380 in the second. In French three colleges had experimental sections and two of them control sections as well, while in German two colleges had experimental sections and two others ran the controls.

Table 17 gives comparisons of achievement and course grades for the first semester of 1961-62. In French no significant differences were found either in the within - college comparisons for College II and for College III, or when the results of laboratory instruction at College I is compared with merged information from non-laboratory sections at Colleges I and II. In German, on the other hand, where both sets of results involve different colleges as well as different laboratory situations, statistically significant results were recorded in reading proficiency tests scores and -- in one instance -- in course grades. The two reading test comparisons are opposite in direction, one favoring the laboratory situation, the other the non-laboratory setting. The differences, although statistically significant, are small and it seems highly likely that they and the course grade differences between College VI and College III are related to other differences between the colleges involved rather than to the language laboratory situations.

The results of testing at the end of 1961-62 for col-

TABLE 17

AVERAGE TEST SCORES AND COURSE GRADES BY LANGUAGE FOR GROUPS USING LANGUAGE LABORATORIES COMPARED WITH THOSE NOT USING LABORATORIES

First Semester, 1961-62

Language and Procedure	Speaking		Listening		Reading		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean
French								
Laboratory (College I)	57	48	79	45	79	36	79	3.07
No laboratory (Colleges II & III)	58	49	83	46	81	36	83	3.28
Significance		NS		NS		NS		NS
Laboratory (College II)	56	51	60	45	59	36	60	3.46
No laboratory	85	50	87	45	88	36	88	3.40
Significance		NS		NS		NS		NS
Laboratory (College III)	—	—	24	46	25	39	27	2.80
No laboratory	—	—	23	46	22	38	23	2.88
Significance				NS		NS		NS
German								
Laboratory (College IV)	—	—	20	44	20	40	20	3.35
No laboratory (College V)	—	—	17	42	18	41	18	3.51
Significance				NS		.01		NS
Laboratory (College VI)	—	—	86	44	80	40	88	3.37
No laboratory (College III)	—	—	68	42	68	39	68	3.09
Significance				NS		.01		.01

leges participating in the lab versus no-lab experiment are summarized in Table 18. In this table, data from different colleges are pooled. In French, the statistically significant differences (in listening comprehension and course grades) favor the non-laboratory group. In German, the test scores all favor the laboratory with differences in mean scores in listening comprehension and reading -- as well as course grades -- that are highly significant from a statistical point of view. The conclusion that language laboratories are effective in German but not in French is highly dubious. One cannot escape the inference that other factors are determining these differences.

Experiment with Various Types of Laboratory Equipment

By 1962-63 all of the member colleges had language laboratories and the emphasis in experimentation naturally shifted away from the question of the use of the laboratory versus no-laboratory to a consideration of the effectiveness of different types of laboratory installation. The colleges with laboratories had begun a comparison of different types of laboratory installations in 1961-62. Potentially such an inquiry was of considerable financial interest

TABLE 18

AVERAGE TEST SCORES AND COURSE GRADES BY LANGUAGE FOR ALL COLLEGES PARTICIPATING IN COMPARISON OF LANGUAGE LABORATORY SITUATION WITH NO LABORATORY

YEAR AND LANGUAGE	SPEAKING		LISTENING		READING		GRADES	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean
French								
Laboratory	105	49	129	38	134	38	139	3.26
No laboratory	85	50	92	40	91	38	93	3.52
Significance		NS		.05		NS		.05
German								
Laboratory	55	51	99	47	94	47	94	3.51
No laboratory	57	50	58	41	57	43	58	2.94
Significance		NS		.01		.01		.01

since it involved contrasts between laboratories ranging in cost from $170 to $750 per student position. An intriguing question is the extent to which these differences in costs are reflected in significant differences in student progress in learning a second language.

The purpose of this experimental design was "to determine the relative effectiveness of five different types of language learning activities as supported by significantly different types of electro-mechanical devices or laboratory installations." The laboratories were divided into four distinct types, the basis for classification being the degrees of individualization of control by the students and flexibility of management by the teachers.

The four types of laboratory set-ups were described as follows:

> Type 1: listening - repeating without recording. No earphones or activitated mikes. Students listen to material from tape recorder or phonograph speaker.

Type 2: listening - repeating without recording (except for selected students when lab is monitored). Activated earphones and mikes at each station.

Type 3: listening - repeating - recording - evaluating by remotely controlled continuous loop-device. Student does not control playback but must wait for loop to play its cycle in order to listen again to the program.

Type 4: listening - repeating - recording - evaluating with completely flexible, individual dual-channel twin amplifier tape deck at each station.

Participating in the experiment in 1961-62 were eight colleges, with sections in French, German and Spanish. Their equipment and its utilization fell into the following classifications:

Type 1: one college (2 sections of French, 3 of German, 2 of Spanish)

Type 2: Four colleges (8 sections of French, 5 of German, 8 of Spanish)

Type 3: one college (3 sections of French, 2 of German, 4 of Spanish)

Type 4: two colleges (6 sections of French, 3 of German, 1 of Spanish)

Each of the three languages, therefore, was taught in each of the four types. When the experiment was continued into 1962-63, however, the types under study were reduced from four to two, and the single-college types were dropped, i.e. Type 1 was merged with Type 2 and Type 3 was merged with Type 4.

In considering the results of this series of experiments it is important to bear in mind that the comparisons include much more than differences between types of language laboratories. Entering into the comparisons are also the differences between colleges with all that this implies with respect to differences in academic environments. Witn this in mind the variations in results turned out to be remarkably small. The major conclusion is that as those laboratories were used in 1961-62 and 1962-63 there is no clear-cut evidence that one type is superior to another. In the following discussion, the results will be organized in terms of the laboratory types, but the complexity of the forces that bear on these outcomes should be kept constantly in mind.

Table 19 shows the mean achievement scores and course grades for the different laboratory types at the end of the first semester of 1961-62.

In spite of the fact that there were three statistically significant differences on the French tests the actual range of means was quite small. The Type 1 laboratory students scored highest on the speaking test and the Type 2 laboratory students had the highest mean scores in reading. There were no statistically significant differences in the German or Spanish tests. The less elaborate laboratories come out very well in these comparisons.

The results of testing at the end of 1961-62 are presented in Table 20. In French the Type 3 laboratory students had the highest mean score on the speaking test, but Type 2 laboratory students were at the top in listening comprehension and reading ability. In German the Type 4 laboratory students excelled on the speaking test and tied with the Type 2 laboratory students for the top spot in listening comprehension. The highest mean score on the reading test was recorded by students from the college with

TABLE 19

AVERAGE ACHIEVEMENT SCORES AND COURSE GRADES BY LANGUAGE FOR ALL COLLEGES PARTICIPATING: COMPARISON OF DIFFERENT TYPES OF LABORATORY INSTALLATION

First Semester, 1961-62

Laboratory Type	Speaking No.	Speaking Mean	Listening No.	Listening Mean	Reading No.	Reading Mean	Grades No.	Grades Mean
French								
Type 1	56	51	60	45	59	36	60	3.4
Type 2	127	49	160	46	168	37	177	3.4
Type 3	60	50	71	45	70	36	77	3.0
Type 4	85	48	119	45	123	36	125	3.1
Significance		.01		.01		.01		—
German								
Type 1	76	50	79	43	79	39	79	3.4
Type 2	98	50	115	43	108	39	117	3.4
Type 3	23	48	38	44	41	39	48	2.5
Type 4	24	50	65	43	66	39	66	3.7
Significance		NS		NS		NS		—
Spanish								
Type 1	36	49	36	41	36	34	37	3.7
Type 2	72	49	138	41	132	33	139	3.2
Type 3	106	49	110	39	111	33	111	3.0
Type 4	6	50	20	40	21	32	20	3.4
Significance		NS		NS		NS		—

TABLE 20

AVERAGE ACHIEVEMENT SCORES AND COURSE GRADES BY LANGUAGE FOR ALL COLLEGES PARTICIPATING:

COMPARISON OF DIFFERENT TYPES OF LABORATORY INSTALLATION

Second Semester, 1961-62

LANGUAGE and LABORATORY TYPE	SPEAKING		LISTENING		READING		GRADES	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean
French								
Type 1	54	50	51	39	53	36	54	3.54
Type 2	22	50	94	42	95	42	101	3.56
Type 3	48	52	50	39	51	38	51	3.20
Type 4	40	48	90	40	89	40	94	3.30
Significance		.01		.01		.01		.05
German								
Type 1	74	50	78	44	52	36	75	3.20
Type 2	69	49	64	46	100	44	104	3.45
Type 3	13	52	17	44	17	46	19	3.20
Type 4	8	55	57	46	56	45	53	3.77
Significance		.01		.05		.01		.01
Spanish								
Type 1	36	49	37	43	36	39	37	3.54
Type 2	57	50	103	44	101	38	102	3.16
Type 3	77	50	82	43	82	38	76	3.03
Type 4	4	52	14	45	13	41	14	3.56
Significance		NS		.05		NS		.05

the Type 3 laboratory. Only one statistically significant comparison appeared in Spanish. The students from Type 4 laboratories had the highest mean scores in listening comprehension.

By 1962-63 the comparisons were reduced to only two laboratory types -- non-recording laboratories (Type 2) and laboratories in which the students' utterances were recorded (Type 4). As may be seen from Table 21, the students in French from the Type 4 laboratories outperformed those from Type 2 laboratories on all the tests. In the other two languages the Type 4 students were superior on the speaking test, but not in reading or writing. In the Spring of 1963 the speaking tests were administered by recording the students' performance. One may wonder whether the results on these tests were affected by the fact that the students at colleges with Type 4 laboratories had more experience with recording in the foreign language.

TABLE 21

AVERAGE ACHIEVEMENT SCORES AND COURSE GRADES BY LANGUAGE FOR ALL COLLEGES PARTICIPATING: COMPARISON OF DIFFERENT TYPES OF LABORATORY INSTALLATION

1962 -63

Year and Language	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
French										
Type 2	77	50	83	26	85	32	86	64	86	3.56
Type 4	28	56	36	32	36	38	15	71	36	3.44
Significance		.05		.01		.01		NS		NS
German										
Type 2	40	50	40	22	40	28	—	—	40	3.02
Type 4	15	63	17	25	16	25	—	—	18	3.61
Significance		.01		NS		.05				.01
Spanish										
Type 2	36	46	38	21	38	31	38	84	37	3.00
Type 4	51	52	54	22	57	24	40	77	58	3.84
Significance		.01		NS		.01		.05		.01

V. INDIVIDUALIZATION OF INSTRUCTION: THE PLACEMENT OF LANGUAGE STUDENTS

Problems of Placement

One of the most difficult problems faced by language instructors is the variety of linguistic abilities among the students in most foreign language classes. When during a class session, the instructor is working with the ablest students in the class, the proceedings may be quite unintelligible to the least able. When he directs his attention to the difficulties that beset the poorer students, the best students may be bored. Some instructors develop great skill in dealing with this dilemma, but in almost every case the rate of progress for the class is a compromise between the rate most appropriate for the linguistically gifted members of the group and the rate which would be most effective for those of more limited ability for language learning. The ideal solution would be instructional procedures that would challenge each student to proceed at the maximum rate feasible given his aptitude for language learning. This, however, is virtually impossible with class rather than individual instruction and the larger the class the greater the difficulties are likely to be.

The problems of the appropriate level and pace of work for a language class are compounded when differing previous instruction has produced varying degrees of proficiency in the foreign language. The students of highest achievement may find themselves repeating material already mastered. This is not necessarily bad, since most students may improve their fluency and skill even when working with familiar materials. But there is a danger that these students will find it possible to hold their own in class with little effort, will delay their progress in learning the language and will develop attitudes and habits that will be inimical to optimum progress when a more advanced stage is reached. The student at the other end of the scale, who finds that the course content and pace assume a greater mastery of the language than he has attained, may fail to master fully the material at a given stage before the class moves on to the next stage. He may thus proceed, dragging an ever lengthening chain of ignorance behind him until, overcome by the frustration, he becomes convinced that he is incapable of learning a foreign language.

It might appear that it would be desirable to group students just starting a language in accordance with linguistic ability, and thus have a "track" system in which the range of abilities in a class could be reduced to an inconsequential level. There are a number of reasons why such a procedure is not feasible in most liberal arts colleges. Many language instructors doubt the feasibility of distinguishing language aptitude in advance with sufficient accuracy to justify placement in a "track" system. Reassignment of students after instruction has proceeded for some time is often difficult for both pedagogical and administrative reasons. There is also a built-in barrier to "track" systems in college instruction due to the nature of the academic accounting system. This system is based upon the academic currency, the unit of which is the credit-hour (a semester-hour or quarter-hour as the case may be). This system treats a credit-hour earned in one course as the equivalent of a credit-hour earned in any other course, tenuous as the grounds for such assumed commensurability may be. The presumptions underlying these procedures produce practically insoluble problems for a "track" system. Should the students in the different "tracks" receive different amounts of credit, or different average grades?

The situation with respect to students who have already had some language instruction is somewhat different. It should be more feasible to discover proficiency achieved than aptitude for language learning in advance of language study. Courses at different levels will normally be offered, and there are thus available alternative opportunities for placement which may make possible a rough correspondence between the actual level of achievement of the student and the average achievement of the class to which he is assigned. In most cases this correspondence will be rough because typically there will be a difference of one year of college study between the courses being offered simultaneously. Even if the student delays starting his college language study in order to achieve closer correspondence, the alternatives available will represent differences equivalent to a term of college study. The placement problem normally faced, therefore, is whether to place the student in a class that seems somewhat advanced in terms of his estimated proficiency, or in a class which, on the average, is below him in proficiency.

In some cases academic accounting poses an additional difficulty for appropriate placement. Some colleges consider that students have, in effect, exchanged their high school credits for admission. Consequently, to give a student college credit for completion of a course in college that is considered to be of the same level as a high school course he offers for admission is to make double payment for his efforts. An example of this position is the so-called "remedial" courses in English, or occasionally in mathematics, which normally carry no credit at all. In contrast to the elementary English and Math taught in remedial courses, all language training is considered appropriate college work, and a college would be very reluctant to deny credit for any college language courses. The usual rate of exchange in language is to consider two years of high school study to be the equivalent of one year of study at the college level. Beloit College will give credit for the course taken in college to a student who is moved back one semester by this reckoning, but not if he is placed back more than one semester. Knox College will give a student credit for the elementary course only if he has had no more than one year of high school study of the language. Other colleges will give a student credit for language study in college if the student has been assigned to a class on the basis of regular placement procedures regardless of the number of years of previous language study.

Methods of Placement

All of the ACM colleges have provisions for placing students at levels other than those determined by the presumed equivalence of two years of high school to one year of college study. During the period covered by this study all of the colleges were concerned with the problem of appropriate placements and a number of changes in procedure were introduced in the effort to discover more effective means of dealing with it. The number of special placements varies considerably from college to college and a number of different methods are used to determine appropriate placement. Very few colleges rely on a single placement device. The nine most commonly used procedures are described below:

1) Grades received in high school language courses

may suggest that the student should be considered for special placement. Unless the high school has an unusually strong program, any grade lower than C raises serious question with respect to the suitability of placement in accordance with the usual relationship between high school and college courses. Students with very high grades from schools with especially strong language programs and three or more years of high school study of a language may be considered for placement in a course in which the emphasis is on literature rather than on language as such, while other students with the same number of years of high school study may be placed in a language course in which the emphasis is more directly on language skills.

2) All ACM colleges require applicants for admission to submit scores on the Scholastic Aptitude Test of the College Entrance Examination Board. Five of them require applicants to take CEEB Achievement Tests. These are one hour tests, three of which may be taken in an afternoon. Four of the eleven achievement tests offered are in foreign languages, French, German, Latin and Spanish.

The CEEB language achievement tests are designed to determine primarily reading comprehension and the understanding of language structure. Thirty-minute listening comprehension tests in the three modern languages have also been made available in recent years.

An indication of the relation between number of years of high school language study and scores on the CEEB Achievement Tests is given by the following tabular presentation:

	AVERAGE COLLEGE BOARD SCORES		
YEARS OF LANGUAGE STUDY	FRENCH	GERMAN	SPANISH
2	421	436	442
3	506	519	519
4	581	554	581*

The College Board Tests are administered five times a year, in January, March, May, August and December. Nine of the ten ACM colleges prefer that candidates for admission take the SAT in December, but three of the five colleges requiring achievement tests prefer that these tests be taken in March.

---

* A Description of the College Board Achievement Tests, College Entrance Examination Board, 1962. pp. 49-50.

There are obvious difficulties in using tests taken in December or March for placing students in classes in the following September.

3) The College Board Placement tests are CEEB Achievement Tests from earlier years that are made available to the colleges for administration by the College, usually during Freshman Orientation Week in the fall. Both language reading tests (which are one-hour objective tests) and listening comprehension tests are available.

4) For a number of years the Cooperative Test Division of the Educational Testing Service published tests at two levels in French and Spanish. The elementary level tests were designed to measure the proficiency achieved after two years of high school or one year of college language study. The advanced level tests were designed for students who had had more than this amount of instruction. The basic tests required forty minutes and tested reading, vocabulary and grammar. A French listening comprehension test requiring thirty minutes was also available.

5) Starting in 1964 the Cooperative Test Division of the Educational Testing Service is offering a new series of tests which have been developed as a cooperative project of the Modern Language Association of America, the Educational Testing Service and the United States Office of Education. They are available in five languages — French, German, Italian, Russian and Spanish. The tests measure reading, writing, listening and speaking. There are two levels. The lower-level tests are suitable for students with two years of high school language, and the higher level tests are for students with four semesters of college work or its equivalent. The time required for the administration of the listening comprehension test is 30 minutes, for the speaking test is 14 minutes, and for the reading and writing tests is 40 minutes each. These tests are published under the title "MLA Cooperative Foreign Language Tests." The lower level tests were used as post-tests in the final year of this program. (See Chapter III).

6) Advanced Placement Tests plus recommendation from teachers in advanced placement courses in high schools may provide the basis for placing students in advanced courses and for granting college credit for work done in high school. These Advanced Placement Tests are three-hour tests and are given at various centers in the spring of each year under the auspices of the College Entrance Examination Board. The examinations test reading comprehension, listening comprehension and writing in the foreign language. They include in varying degrees from language to language literary interpretation and knowledge of literature.

7) The Proficiency Tests for Teachers and Advanced Students were the result of a cooperative project involving the Modern Language Association of America, the Educational Testing Service and the United States Office of Education. They are designed to determine achievement at an advanced level in seven areas, listening comprehension, speaking, reading, writing, language analysis, knowledge of cultural backgrounds and professional preparation as a teacher. As placement tests, this battery is most useful for advanced

students. They require a total of approximately four hours for the full battery. The basic skills can be tested in two hours. The Listening Comprehension Test is received from master tapes with the examinees' responses recorded on answer sheets. The Speaking Test is also received from master tapes but with the examinees' responses recorded on separate tapes or discs. This test was a basic part of the testing aspect of this project. (See Chapter III).

8) Because of the sampling problem discussed in the chapter on the testing program (Chapter III) some language teachers feel that tests designed for widespread administration are less than ideal for placement purposes. Such tests are designed to measure the achievement levels of students who have taken a wide variety of courses. The emphasis and sampling that are inevitable in test construction may not fit well with the preparation the student has had. In placement the courses given at the college are normally appropriate background for the courses at the next higher level. A test based on the elementary course would presumably be best calculated, therefore, to predict

success in the more advanced course. For this reason some colleges prefer to prepare their own placement tests designed to predict where the student who has studied elsewhere may most happily join the company of those who have had their earlier language instruction at the college in question. Some departments in at least three ACM colleges follow this practice.

9) Most of the language departments in ACM colleges will change the placement of a student if it is apparent during the first few weeks of a course that he does not have the needed background. Many teachers feel that this procedure is more effective than placement tests. Analysis of the difficulties students are having early in the course may make it possible to distinguish deficiencies that are limited and specific (and thus possibly amenable to special remedial work) and those that are general and thus call for a revised placement decision. Students who have been placed at too high a level are more likely to be detected than those who are placed below the most appropriate level. (The latter type of inappropriate placement is quite unusual when the standard formula for equivalence between

high school and college courses is used.) The handicap of absence from the class for the initial period is also greater for the student who is shifted up than for the one who is shifted down.

ACM College Procedures in Placement

There is clearly a considerable variety in the procedures used for placing students in the appropriate college language course, and serious efforts are being made to develop better procedures. Faculty members concerned with proper placement have encountered a number of difficulties, some inherent in the instruments used, some resulting from differences in pedagogical objectives and techniques and some from administrative complications.

One difficulty with the use of high school grades is the great variety of standards used for grading, not only between schools but between teachers in the same school. If a college draws large numbers of students from particular high schools, it may be possible to acquire a fairly accurate sense of the meaning of the grades given in these schools. Where the college draws its students from a large number of high schools with few from each, there is a strong tendency to assume that a given grade has the same meaning regardless of the school from which it was received. In spite of these deficiencies a study of the grades received in high school may be useful and may be a means of discovering situations which should be subject to further investigation.

One important limitation on the use of high school grades and a significant problem in the articulation between school and college language study is the lapse of time that frequently occurs between study at the two levels. In spite of improvements in recent years, the typical high school graduate has had only two years of language study and this is frequently taken in his first two years in high school. The atrophy in his language competence by the time he enters college is normally so great that he cannot hope to fit into a second year college class geared to the competence of students who have just completed a year of college study of the language. Where there has been a hiatus of one or two years in language study it may be possible from grades to determine which students should be placed at the elementary level.

Time discrepancies are also significant in the use for placement of tests given during the senior year in high school. CEEB Achievement Tests taken in December-- or even in March--can hardly be an accurate indication of the language competence with which students come to college in the fall. On the other hand, tests taken in May or June may be quite satisfactory, since the students in the college courses will not normally continue their language study during the summer.

A very common practice has been to administer language placement tests during Freshman Orientation Week just before classes begin in the fall. The most common tests used for this purpose are the College Board Placement Tests, (3) above, the Old Co-op Tests, (4) above, and the New Cooperative Tests, (5) above. For colleges in which there is emphasis on listening comprehension and speaking, only the new Cooperative Classroom Tests are suitable.

There are a number of problems that make placement testing during Orientation Week less than ideal. At some colleges students are subjected to a massive testing program during this period, and their performance may be adversely affected by fatigue or resentment. They may also consciously malinger in a test which is used for placement purposes in the effort to achieve an assignment that will make it possible for them to obtain good grades with little work.

The task of grading large numbers of tests taken during Orientation Week in time for placement of students is formidable. This is particularly true of a test designed to evaluate speaking ability. The process of assigning students to courses just as classes are beginning may entail many administrative difficulties, including rescheduling of classes and rearrangement of teaching schedules.

The reactions to these difficulties have been in two directions. One is to rely more heavily on high school records and tests administered before the student comes to college. These are the CEEB Achievement Tests, (2) above, and the Advanced Placement Tests, (6) above. There is some interest in having the new Cooperative Classroom Tests administered in the spring, presumably at high schools, so the results can be available for placement decisions to be made during the summer.

The other direction in which colleges have moved is to change placements on the basis of class-room performance during the first week or two of classes. To be effective such a procedure requires heavy emphasis on diagnosis at the outset of the course. Administrative difficulties may discourage large scale reassignments after classes have started. As indicated earlier, the shifts are more likely to be downward than upward, and students capable of more advanced work may be retarded in their progress.

Placement Results

At five of the colleges included in this study, records of placements in recent years make it possible to observe the results of the process. Table 22 combines the data for all languages and presents the information for each college for the years indicated. This gives the placement outcome for 1,744 students, of whom 729, or 41.8 per cent, were placed in accordance with the presumed equivalence of two years of high school and one year of college language study. A total of 556, or 31.9 per cent, were set back one semester and 319, or 18.3 per

TABLE 22

RESULTS OF PLACEMENT PROCESS AT FIVE COLLEGES

All Languages

	Cornell ('63)		Grinnell ('59 - '63)		Knox ('59 - '63)		Monmouth ('59 - '63)		Ripon ('61 - '63)		TOTAL	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Back 4 semesters	—	—	—	—	—	—	—	—	2	1.5	2	0.1
Back 3 semesters	—	—	—	—	5	0.9	—	—	3	2.2	8	0.5
Back 2 semesters	22	6.4	9	4.1	84	15.3	157	31.9	47	34.3	319	18.3
Back 1 semester	44	12.8	51	23.0	302	55.0	136	27.6	23	16.8	556	31.9
Equivalent level	269	78.2	148	66.6	128	23.3	145	29.5	39	28.5	729	41.8
Ahead 1 semester	4	1.2	12	5.4	19	3.5	21	4.3	13	9.5	69	4.0
Ahead 2 semesters	5	1.4	2	0.9	7	1.3	33	6.7	8	5.8	55	3.1
Ahead 3 semesters	—	—	—	—	3	0.5	—	—	1	0.7	4	0.2
Ahead 4 semesters	—	—	—	—	1	0.2	—	—	1	0.7	2	0.1
	344	100.0	222	100.0	549	100.0	492	100.0	137	100.0	1744	100.0

cent, were set back two semesters. These were the major placement actions, but it may be surprising to note that 69 students were set forward one semester and 55 were set forward two semesters. The typical situation in which a student might be set forward one semester would be one in which he had only one year of high school study of the language but was placed in the second year at college. The typical student set two semesters forward would be one with two years of high school study who is placed in the third year college course. Placements deviating more than two semesters in either direction obviously represent special cases.

The figures in Table 22 reveal some very striking differences in placement practices between colleges. When attention is turned to the information by college, it is seen that the majority of the students at Cornell and Grinnell were placed according to the equivalence presumption, while a minority of the students were so placed at Knox, Monmouth and Ripon. The percentages are 78.2 and 66.6 respectively for Cornell and Grinnell and 23.3, 29.5

and 28.5 for Knox, Monmouth and Ripon. There is another contrast. At Cornell, Grinnell and Knox most of the students who were placed at a lower level were set back one semester. At Monmouth and Ripon the retardation amounted to two semesters for a majority of those set back. No immediate explanation for the first contrast appears. A possible explanation for the second applies to Ripon and not to Monmouth. It has not been the practice at Ripon to offer second semester sections in the first semester. Students who have had two years of high school language training and who are starting college language study at the beginning of the academic year will be placed in the beginning course if their proficiency is not adequate for the second year course. Monmouth, however, has followed the practice of offering second semester sections in the first semester.

Three languages, French, German, and Spanish, enter into the tabulations by college in different proportions. The data for Grinnell and Ripon, for example, are for German only, while the data for the other colleges apply to all languages. The results for Grinnell and Ripon do not suggest that placement patterns differ by language, since the percentage placed according to the equivalence presumption is 66.6 at Grinnell and is 28.5 at Ripon. In Table 23 the results of placement by language are presented. It can be seen at a glance that the patterns are generally consistent for all languages at a given college. An attempt has been made to account for these differences, but no explanation has been found.

Articulation between High School and College Language Study

Another matter of interest is to inquire whether improved language instruction in secondary schools is leading to the placement of increasing numbers of students at the equivalent or more advanced level in College. For three of the colleges placement data were available for a five year period, 1959 to 1963 inclusive, and at one college for the last three years. The data are given in Table 24 and are graphically depicted in Chart 4. There is no consistent evidence of a trend in any direction, but the cases where declining numbers of students are

TABLE 23

Percentage of Students Placed At Equivalence Levels By Language and by College

	FRENCH	GERMAN	SPANISH
Cornell	80.8	78.5	73.8
Grinnell	—	66.6	—
Knox	24.8	20.6	23.6
Monmouth	29.6	40.7	26.8
Ripon	—	28.5	—
St. Olaf	79.4	86.6	65.9

TABLE 24

PERCENTAGE OF STUDENTS PLACED AT EQUIVALENCE OR MORE ADVANCED LEVEL, BY LANGUAGE, SELECTED COLLEGES 1959 - 1963

	1959	1960	1961	1962	1963
Grinnell					
German	63	73	73	80	71
Knox					
French	31	29	29	29	31
German	28	23	22	32	27
Spanish	47	23	36	23	22
Monmouth					
French	46	37	50	44	34
German	67	42	58	43	30
Spanish	44	27	33	49	45
Ripon					
German			46	45	45

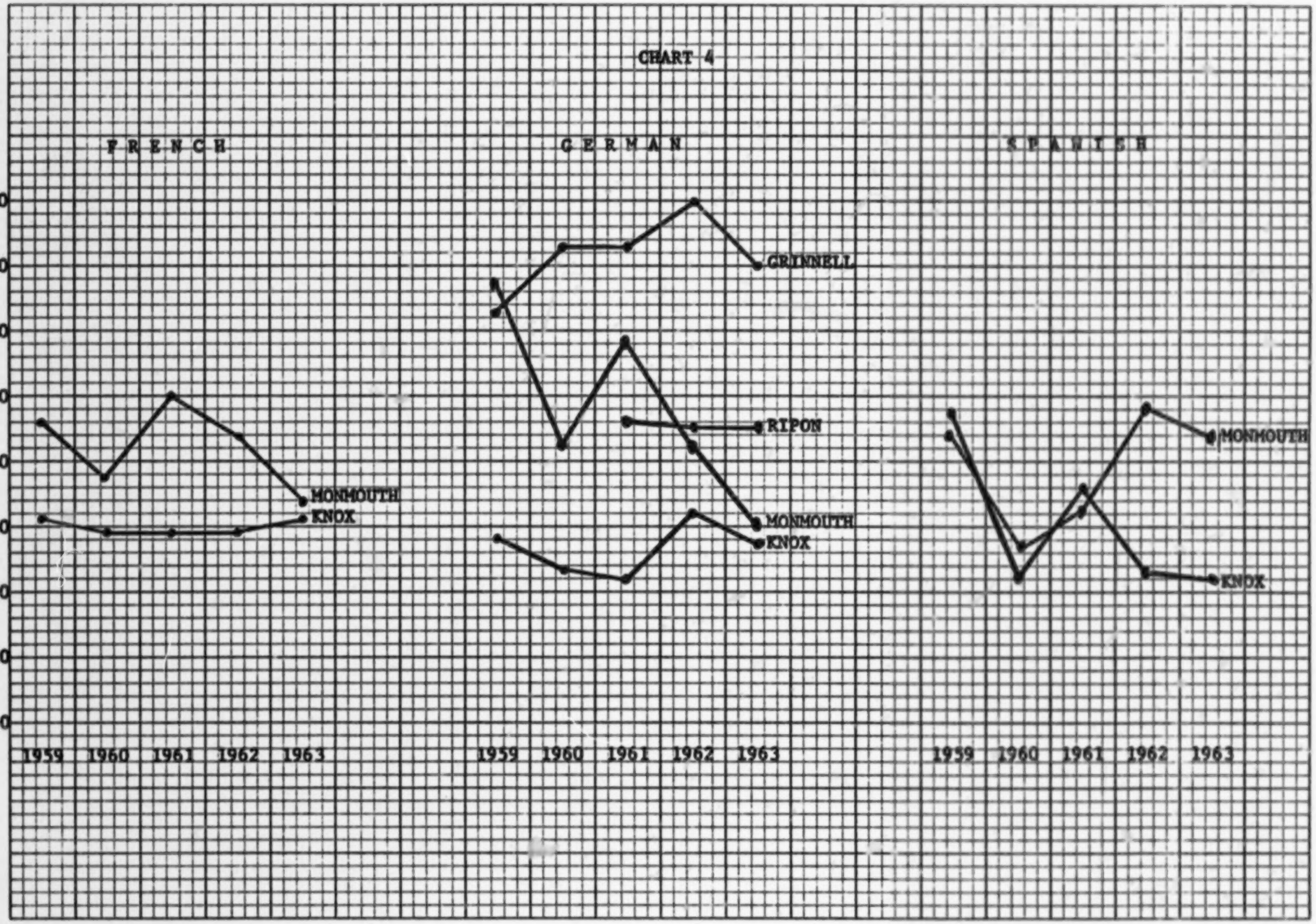


CHART 4

FRENCH
GERMAN
SPANISH
MONMOUTH
KNOX
GRINNELL
RIPON
MONMOUTH
KNOX
MONMOUTH
KNOX
1959
1960
1961
1962
1963
1959
1960
1961
1962
1963
1959
1960
1961
1962
1963

being placed at these levels seem at least as common as those where the number so placed is increasing. Other factors than improvement in previous language training must be sought to explain the changes shown in this chart.

There is one problem in the articulation of language study between high school and college that has an important bearing on placement in college. This is the lapse of time between high school and college language study. As has been noted, it is not at all uncommon for a student to take only two years of language study and this in his first two years in high school. If such a student does nothing further with the language before entering college he usually has a very small residue of the foreign language left. He will make a poor showing on the placement examination and, in all likelihood will be placed in the first semester of the elementary course. He may decide to start another language.

In some cases there are good reasons for a shift in the foreign language studied. In others this shift may be induced by the college policy of reduced credit for students who are placed below the presumed equivalence level. Students who change the language studied for such reasons frequently meet the college foreign language requirement without achieving a lasting command of the second foreign language studied. They have put in four years of language study--two in high school and two in college-- and graduate from college with only a smattering of two languages. A concerted effort at both levels should be made to avoid such a result. Students should, if possible, have four years in language study in high school and should strengthen the competence so gained in college. Where this is not possible, means should be formed to make it possible for students to enter college with their foreign language competence at the peak level they have been able to achieve.

The effective placement of students in college courses is a significant problem for the college, and the colleges included in this study are continuously concerned with it. No procedure that is universally satisfactory has been found, and as long as class instruction continues along present lines the discovery of such a panacea is unlikely. It is possible, however, that increasing individualization of instruction in college may make it possible for students to start where they are in language competence and proceed in accordance with their ability. One possible approach to such a situation--programmed learning--is considered at a later point in this report.

VI. INDIVIDUALIZATION OF INSTRUCTION AND THE LANGUAGE LABORATORY

Conditions of ACM Program

At this stage in the discussion of factors in the general improvement of the quality and efficiency of language instruction, it may be appropriate to recapitulate certain significant conditions of this program. All ten of the colleges which constitute the membership of the Associated Colleges of the Midwest participated, but the selection of areas of investigation by each institution was voluntary, and to a large extent the character of any experiment conducted on a given campus took on an individual coloration from the local conditions and personnel. There was no machinery for imposing procedures, beyond consensus and agreement, and no central authority for dictating consistent and uniform detail.

This voluntary situation further underlined the fact that the investigations under way were not experiments in the laboratory sense but field experimentation, what is called "point-of-application evaluation." Therefore it would be unrealistic to judge relatively soft statistical evaluations by the criteria of the laboratory, or to expect clear-cut and definitive statistical significance.

Field experimentation of this type however is not to be dismissed as superficial and without body. The specifics must be tentative; but the over-all significance lies more in the imponderables of changed atmospheres on the various campuses, awakening of the faculties to the nature and impact of new approaches, and resultant gradual modifications of curriculum and teaching techniques. It should be emphasized that the stated purpose of the program was not the production of watertight statistical evaluations produced under laboratory controls, but rather improving the quality of language instruction. Such a quest for improved quality obviously pre-supposes some examination of new approaches for their relevance and scope, but the more pervasive effect sought would have been compromised had too great insistence been given to precision of controls and rigidity of procedure.

Individualization as a Central Theme

A central unifying theme which emerged throughout the years of the program was the desirability of the individualization of instruction, and in some respects all of the experimental designs were radial to this hub. For instance, several aspects of the value and best utilization of the language laboratory were investigated. The basic impact of the use of a language laboratory was tested by having experimental groups work with a laboratory, and control groups without such availability. Combinations of varying class size and degrees of laboratory supplementation were also examined. In seeking the most effective use of the language laboratory itself, one aspect of the problem aroused interest in the first Summer Workshop in 1960 which was maintained throughout the three years of experimentation. This segment of the problem was the role of a monitor, or put in terms of a comparison of procedures: the effects of monitoring versus non-monitoring of student work in laboratory periods.

The Objectives of Laboratory Use

Such a comparison must of course be evaluated against a definition of what the work in the laboratory is supposed to accomplish, and the spectrum of positions on this point is wide. Some teachers of language still hold that the usual two-year sequence in a modern language should be centered on reading and literature, since it is unrealistic to aim for much more in the time available. Even these, however, recognize the necessity for the drill and repetition which are called for if the student is to master such things as irregular verb forms and they see in the laboratory a device for accomplishing this purpose with more flexibility in scheduling for the student and with a tremendous saving of the teacher's time and energy. Such a limited intention in the use of the laboratory obviously does not envisage much monitoring.

At the other end of the spectrum is the advocate of the audio-lingual approach in all its purity. Here the objective is to establish in the student, as "patterned human social behavior", a set of automatic psycho-motor responses which can only be built up by habit. Obviously, here there must be safeguards which prevent the student from establishing responses which are in themselves inaccurate. Carried one step further, where the intention is conceived to be the closest possible approximation to the sounds native to the language being studied, detailed supervision and correction are called for. This position also tends to hold that the student should from the beginning hear the sounds properly made, and should be immediately corrected when in his attempts to approximate the sounds he deviates perceptibly, even though he himself may not recognize the difference. In such a case quite close monitoring and intervention by the monitor would probably be a natural result.

Whatever the definition of objective, there are four obvious increments from the use of the language laboratory: 1) the opportunity for the necessary drill and repetition, of whatever variety decided upon; 2) the opportunity to hear a native speaker from tapes when such an individual is not available in person on the particular campus; 3) the opportunity for the student to adjust his own schedule to some extent, and to move at his own speed (individualization of instruction); and 4) a beneficial change in the teacher's role by releasing him from the supervision of repetition for its own sake, thus freeing him to devote his time and energy to the more constructive and creative aspects of language teaching. (An interesting variation involving use of the language laboratory which integrates it very intimately into the classroom procedures is developed in the Experimental Material in German described in Chapter VII on programmed learning.)

The nature of the objective determined upon within each of these possibilities, and the balance among the four, will of course be a function of the philosophy and procedures characteristic of the faculty, and even of individual teachers, and will in turn determine the configuration of the pattern of utilization followed.

Monitoring versus Non-Monitoring

The original experimental design in the ACM Language Program made a very definite differentiation between monitoring and non-monitoring. "Monitoring" was defined as meaning that the course instructor listens to the student's lab practice, making corrections as necessary; while "non-monitoring" was defined as having some person other than the course instructor operate the lab without listening to individual students and without correcting them.

As a result of the first year's experience, the experimental design used in both 1961-62 and 1962-63 pointed up the need for greater specificity as to the different kinds of monitoring to be used. Is monitoring simply a pedagogical device to assure that all students are actively doing the prescribed laboratory work? Should the monitor be a language teacher who can correct student errors? How much interruption for correction should be engaged in? Should corrections be made only to the erring student or to the entire class?

The experimental design followed these questions with three proposed varieties of monitoring:

a) An instructor of the language merely listens to performance of individuals in lab work.

b) The instructor may, in addition to listening, correct and coach the student individually.

c) The instructor may furnish the entire group with the error and its correction. Normally not more than 25 percent of the lab period will be devoted to this activity.

Non-monitoring was described as student work in the lab as a practice period without supervision or corrections by an instructor. Normally no instructor would be present, and only a student lab assistant would be in attendance. Such an assistant would be instructed only in the operation of the equipment and be informed what materials students were to use in each lab period.

As the daily instruction routine unrolled itself, three varieties of procedure crystallized. In one, the student was supposed to apply himself to prescribed material without supervision; in the next, the student was supervised, usually by another student, with the objective of insuring application to the task in hand, and this procedure was in some cases labeled "monitoring" and in the third, an instructor listened at random but intervened only slightly or not at all, this procedure also being labeled "monitoring."

The rationale for settling on the three procedures outlined can be found in a consensus of the language teachers involved.

In the first place, they agreed on the general thesis that there is a shift in what should be the major emphasis as between beginning language instruction and later phases. At first, the most fruitful learning experience for the student is the greatest possible amount of listening and repeating without intervention, and this type of exercise should be scheduled and controlled (i.e., supervised.) In more advanced stages of language learning, the student profits most by shifting the major emphasis to speaking, with considerably more flexibility in materials used, scheduling, unsupervised practice, etc.

This summary formulation grew out of several convictions based on experience in the program. The faculty group repudiated the argument that a student must be closely monitored and never allowed to make a sound incorrectly without correction lest he harden into incorrect habitual responses. They believed their experience showed that, after anywhere from 50 to 200 repetitions of a cluster of phonemes, a student would himself develop the necessary discrimination and tend to correct himself without intervention by a teacher.

They further believed their experience indicated that the gains from intervention did not compensate for the time taken away from the more meaningful repetition and drill, and the practice was in fact counter-productive in terms of the distraction and the increased tension induced in the student. Their basic thesis might be restated as a conviction that in the early stages of language learning, the habituation of the student to patterns characteristic of the language being studied was far more valuable in terms of fluency than in terms of exactitude of sound approximation. They felt that the approach of never letting the student say it wrong just did not pay off in terms of the student's total progress.

A condition which might have influenced their reaction was the fact that the ACM colleges are so well supplied in general with native speakers of the languages being taught that the opportunity to hear a native speaker through the laboratory became relatively less important, with a consequently increased role for the practice and drill components.

Attention should be drawn to the fact that this consensus was the widely-held position of experienced language teachers. The majority of them were not linguistically trained, nor had they had much contact with language laboratories before the program began, and some more ardent advocates of the audio-lingual approach would undoubtedly cavil at this inexperience; but the fact remains that the reaction of these experienced teachers was remarkably consistent.

Furthermore, notice should be taken of the fact that these conclusions were judgmental in nature, and not susceptible to validation by refined statistical data. Within the open framework in which the whole program was conducted, however, these conclusions seem worthy of serious consideration. At any rate, they contain some concepts which offer interesting fields of investigation through more rigorous experimentation for the skeptically minded.

Before consideration of the statistical data and its possible interpretation, one off-shoot of the individualization of instruction approach in the language laboratory should be mentioned. The most inflexible scheme for student access to material is to have each student scheduled for his entire practice time, and the materials to be made available to him specified. A relaxing of this procedure might enable the student to go to the laboratory in whole or in part at his own convenience, but to specify the amount of time to be spent and the materials to be covered when he does go.

Participation

In the experimental evaluation of monitored vs. non-monitored language laboratories, four colleges participated during the pilot year, 1960-61. Of these, two had language laboratories already installed, while the participation of the other two was made possible by the utilization of four portable units of the inflexible type (i.e., no play-back, recording, etc.). In these four colleges, there were six monitored (experimental) sections in French, six in German, and two in Spanish as against four un-monitored (control) sections in French, four in German, and four in Spanish.

In all, about 600 students who were beginning the study of the respective languages took part.

In the next year (1961-62), six colleges took part, but with a reduced total of sections and students. There were three monitored sections in French, two in German, and five in Spanish, with four un-monitored in French, three in German, and six in Spanish as controls. In the first semester approximately 450 students took part, with about 275 in the second.

Finally, in the last year (1962-63), only three colleges participated. They produced three monitored sections in French, three in German, and two in Spanish, with one un-monitored section in French, one in German, and two in Spanish reported. The number of students involved was approximately 225.

In summary, eight of the ten ACM colleges took part in this study, two of them for all three years, one in the first and third years, and five others for one year only, one of them in the first year and four in the second. The data for the third year thus represents a concentration of experience, since two of the colleges were in their third year of participation, and the other in its second after a year's lapse.

Statistical Analysis

In discussing the statistical indications of the results of the experiment, certain conditions should be borne in mind. The pilot year was considered as largely exploratory, and such statistics as were compiled were not in conformity with the methodology applied to the more solid succeeding years. In the second year, post-tests were administered for both semesters, while in the last year, statistical data were based on only one set of tests at the end of the entire year of experimentation. Also, the tests used during 1962-63 were not the same as those used in 1961-62.

The varying numbers of sections, languages, etc., in the program to a certain extent dictated a different statistical approach from period to period, a tendency reinforced by a desire to examine differing kinds of relationships. In strictly research terms, the tests were somewhat inconclusive, since most of the comparisons did not produce differences sufficiently marked to be statistically significant, and the relatively few such cases were not consistent in their direction.

Of the various tables of data, those giving an analysis of the individual work of one college in one language may be considered as more realistic in many ways, since differences between colleges and between approaches to different languages probably have considerable effects on pooled data. In the first semester of 1961-62 two such studies were made, and in 1962-63 four. One college experimented with Spanish in the first year and German in the subsequent year; one college used French in the second year; and a third college used German in the first year and both Spanish and German in the second. The results are set forth in Table 25.

Table 25 contains only three cases of statistical significance. In 1961-62 the non-monitored students in College I performed significantly better than the monitored students

TABLE 25

POST - TEST RESULTS AND COURSE GRADES FOR MONITORED AND UNMONITORED LABORATORIES

	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
FIRST SEMESTER, 1961 -1962										
College I-Spanish										
Monitoring	17	46	20	40	18	33	—	—	20	2.96
Non-monitoring	29	50	33	42	31	35	—	—	34	3.30
Significance		.01		NS		.10				NS
College II-German										
Monitoring	15	48	22	45	23	40	—	—	28	3.79
Non-monitoring	9	48	16	43	18	39	—	—	20	3.45
Significance		NS		NS		NS				NS
1962-1963										
College I-German										
Monitoring	35	63	38	34	36	32	36	90	39	3.72
Non-monitoring	20	63	20	32	19	32	20	89	20	4.10
Significance		NS		NS		NS		NS		.05
College II-French										
Monitoring	31	53	31	27	29	28	33	50	37	3.35
Non-monitoring	13	56	15	27	11	28	14	48	15	3.67
Significance		NS		NS		NS		NS		NS
College II-Spanish										
Monitoring	21	53	20	19	21	21	20	69	24	2.79
Non-monitoring	23	56	24	20	23	22	24	72	24	3.00
Significance		NS		NS		NS		NS		NS
College III-French										
Monitoring	21	41	23	23	22	34	23	61	23	3.22
Non-monitoring	13	42	15	22	15	36	14	56	15	3.53
Significance		NS		NS		NS		NS		NS

on the speaking test and at a level approaching significance on the reading test in Spanish at the end of the first semester. In the same college the following year the non-monitored students received statistically significant higher average grades in German at the end of the year. In view of the fact that the numerical equivalents of the letter grades are a series of digits summing from 1 to 5, not much importance can be attributed to this difference. No other differences between monitored and non-monitored sections were even statistically significant.

To see whether pooling of data might yield statistically significant results a summary examination was made merging the data for a given language for all colleges and merging the data for the three languages for all colleges. These statistical manipulations produced four statistically significant results (all favoring non-monitored labs) in 1961-62, but none in 1962-63. As noted above, this pooling of data tends to increase the unidentified variables.

One very interesting contrast is shown in Table 25.

This is the contrast in performance on the different tests in French at College II and College III. An emphasis on the audio-lingual skills at College II is reflected in average scores on the speaking test of 53 and 56 for the monitored and un-monitored sections as compared with scores of 41 and 42 at College III. On the listening test the average for both sections was 27 at College II as compared with averages of 23 and 22 at College III. In reading and writing, on the other hand, the students at College III outperformed the students at College II. In reading the students at College II had mean test scores of 21 and 22 as compared with scores of 34 and 36 at College III. On the writing tests College III students averaged 61 and 56 as compared with 50 and 48 at College II.

Conclusions

If the hypothesis with respect to monitoring is stated in the null form, i.e. that monitoring does not contribute significantly to the performance of students in speaking or the other language skills, nothing can be found in this study to disprove the hypothesis. Until more extensive and carefully controlled studies do undermine the hypothesis it is appropriate to doubt whether it is worthwhile to install equipment that is to be used solely for monitoring and to use any substantial amount of the time of faculty members in this way.

The Library Use of the Laboratory

The library use of the language laboratory is a freer type of study situation than regularly scheduled, but unmonitored laboratory periods, and lends itself to greater individualization of instruction. In this setting, the student can go at any hour that suits him, and have access to any material that he feels in his own judgment he should work on. The observation of the language teachers was that the students spent more time, and with better motivation, when the laboratory period was not scheduled; a fixed hour for the drill or outside class practice period seemed to arouse some feelings of resistance. Within the unscheduled framework, however, there was no breakdown which would give a basis of judgment as to whether specified materials or library use seemed to give the best result.

A further extension of the library approach was the "take-home disc" procedure. Here discs are cut from the laboratory tapes to match the material made available to the student in the laboratory. The student who draws the records out to take to his room also draws a record-player with ear phones. An attempt to increase motivation in, and hence utilization by, the students was made by the instructor, in that he pointed out to them they had the chance to work more intensively on the material, and urged them to take full advantage of their opportunity.

The use of discs was investigated only very tentatively, since it was tried out in only one college in the year 1961-62 in French. The adoption of a revised edition of the text disrupted the original dove-tailing of the material on the discs with the classwork; in all probability, the glamor of a new and elaborate language laboratory also dimmed the interest of the faculty in further use of the discs or the cutting of new ones. The members of the department had already agreed that they were somewhat less than enthusiastic, since they were convinced that the students were not giving the proper attention to this unsupervised type of drill. The head of the department felt that the conditions of dormitory life made the use of discs difficult (roommate's objections, distraction from radios, etc.). He further commented that even in the laboratory the students had to be checked quite rigorously to insure their practicing the expected time. Certainly such an attitude on the part of experienced teachers raises a question of how ready the average college student is to profit by entirely unsupervised practice-drill opportunities.

The actual data on the achievement levels of so small a group is naturally inconclusive. Within the college concerned, a non-monitored lab section was considered the control group. This latter class had the same class time (4 periods of 50 minutes each per week) and 4 scheduled, required laboratory periods of 30 minutes each. A student assistant was in attendance only to take roll and play the assigned material. Everyone in the lab heard the same material. The laboratory equipment permitted only listen-respond without recording, with earphones and individual volume controls. The comparative figures of these two sections are given as a matter of interest in Table 26. No clear pattern of results emerged with the possible exception that in both semesters the students practicing with discs did somewhat better on the reading test and received somewhat lower overall grades than the students who made regularly scheduled use of the laboratory.

The discs were offered to other colleges at no cost, but the teaching staffs did not show enough interest in the project to take advantage of the opportunity. Therefore the examination of this particular technique gives only a slight basis for judgment as to its value, and further testing would be desirable. It seems that for the student with a high degree of motivation, such a resource as the library use of the laboratory as well as the take-home disc would offer a real opportunity for greater effectiveness in learning a language. However, in a language course required to fulfill graduation requirements a significant proportion of the students may not be sufficiently motivated to capitalize fully on such opportunities. The library use of the labora-

TABLE 26

POST - TEST SCORES AND COURSE GRADES:

TAKE-HOME DISCS COMPARED WITH NON-MONITORED LABORATORY

	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
First Semester, 1961-62										
Discs	27	48	28	46	27	36	—	—	28	3.36
Laboratory	29	53	32	45	32	35	—	—	32	3.44
Second Semester, 1961-62										
Discs	—	—	23	37	25	37	—	—	24	3.36
Laboratory	—	—	30	40	30	36	—	—	31	3.65

tory and the take-home disc technique may be thus limited in their application, but may have a high potential in the cases where the student is prepared to exploit them fully.

VII ..INDIVIDUALIZATION OF INSTRUCTION: PROGRAMMED LEARNING

Materials and Implementation

The possibility that programmed learning might contribute substantially to foreign language instruction and particularly to instructional programs permitting each student to proceed at his own pace was the subject of discussion at all three of the faculty Summer Workshops. Several persons who had been involved in the production and use of programmed materials for language learning were invited to the Workshops for lectures and group discussions. Marcel Moraud told about the use of programmed materials in French at Hamilton College. Stanley M. Sapon and Maurice W. Sullivan discussed their respective philosophies with respect to programming and described the programs in Spanish that they had prepared. Dale Brethower of the University of Michigan described the programming work done there under the direction of Rand Morton.

There was interest on the part of a number of instructors in starting experimentation with programmed learning as soon as possible but it was not until the fall of 1962 that suitable materials for experimentation with programmed instruction became available. Even then the published programs were unfortunately incomplete. It was decided that the most suitable programs were the Temac materials published by Encyclopaedia Britannica Press. Spanish A by Stanley M. Sapon and German A by Ernest E. Ellert were the materials selected. Both of these are linear programs of the Skinnerian type. They proceed by a series of questions, each of which is answered by the student, who is given the correct response before he proceeds to the next question. Both of the programs involve a workbook and tapes. To go through the programs the students needed to have access to tape recorders with foot pedal controls. The Inland Steel-Ryerson Foundation made a grant which was used to defray part of the costs of providing 59 tape recorders for use in these experiments. Six colleges and eleven faculty members participated in the programmed learning experiment. The Spanish materials were used at five colleges: Beloit, Cornell, Knox, Monmouth and St. Olaf. The German materials were used at three colleges: Grinnell, Monmouth and St. Olaf. The experiment started with 89 students in Spanish and 57 in German. At some colleges the participating students were selected at random from students enrolling for the elementary course and at other colleges volunteers were called for.

There was considerable variation in the way the programmed materials were related to classroom instruction. There were eight instructional groups, five in Spanish and three in German. One group had two class meetings a week in addition to the work on the programmed materials, five met once a week and one met only once every two weeks. In all of the groups the programmed learning was supplemented by a testing program. The number of tests given during the period which the programmed learning materials were used were 2 for one group, 5 for two groups, 10, 11 and 17 for one group each and 1 group was given a short oral quiz each week. In some cases the tests were graded and in others they were used merely as a guide to the student and instructor on student progress and degree of mastery of materials covered. In at least one college, supplementary materials were designed in an attempt to keep the students using programmed procedure in step with the subject matter coverage of the regular sections. This activity was related to the problem of transition discussed in the next paragraph.

Problems of Transition to Conventional Methods

On the average the programmed materials seemed to be suitable for approximately ten weeks of college level instruction. Because the programmed materials differed so much from conventional instructional materials, there was a problem of how to proceed further with the instruction of the students who had participated in this program. This problem was met in a number of different ways at the different schools. The major difficulty was that the available materials were sufficient for only a portion of the first term, and the anticipated material to finish out the period was not forthcoming. A meeting of representatives of the colleges concerned was held to discuss possible solutions, and several variant transitional procedures were developed on the different campuses.

The comments of one instructor in Spanish illustrate the difficulties arising from the absence of material to complete the term using programmed material.

> "The disappointment and frustration of the students was very keen upon learning that no further programmed materials would be available.

> Once the transition to a standard classroom situation was made I found it very difficult to cope with the general collapse in morale. It took a good deal of patience and effort to rebuild their interest."

In College I, a two-week transition period was used to pick up the items which the control section had covered in the regular text but which the experimental section had not had, and then from that point forward the procedures of the regular sections were followed. A manful attempt was made to maintain the audio-lingual emphasis in class and through use of the language lab.

In College II, the entire remainder of the term was organized to bring the students up to the conventional classes in subject matter, while continuing the emphasis on audio-lingual exercises.

The instructor in College III went rapidly through the regular lessons already covered in the conventional classes, pre-selecting the unfamiliar items and giving major emphasis to memorizing dialogue and to oral drill on new points. The same general procedure of selection was continued in the second semester, although even at the end of the year the students were not up to the usual point in subject matter coverage. Notwithstanding this circumstance, the programmed material and follow-up was repeated in 1963-64.

Colleges IV and V tried somewhat more ambitious transition plans in spite of the effort required. To fill out the term, the instructor in German at College IV prepared, as his department head put it, "countless" questions and answers based on the graded readers, which he administered orally (but did not tape). This material was designed to maintain the general approach of programmed learning, and at the same time to fill in the gaps in subject matter coverage. The combination of German A and this locally developed material was repeated in 1963-64.

College V anticipated the problems ahead, and early in the term began to make adjustments. The programmed material in Spanish was spread out over the 15 weeks of the term, and supplemented with a) additional vocabulary of high frequency words, b) some graded reading selections, and c) two class meetings a week for grammatical pattern practice (using only the patterns given in Spanish A) and simple conversations. The fact that the class was small made such a program easier to carry out.

> "During the second semester we did not have any programmed learning materials; therefore we had to switch our students into a regular second semester language class. However, we tried to keep some of the programmed learning techniques in the classroom situation. In order to make this change, and because our students were rather behind in grammar and vocabulary, compared with the other regular classes, we set an intensive and fast pace supplying carefully selected grammatical pattern drills, and a great deal of vocabulary in contextual form, in dialogues and reading selections. Our students were also subjected to weekly short quizzes to test their performance.
>
> "Some of our students were unable to adjust themselves to this new pace, especially the ones who had other motivations than learning."

This same College V, although expressing some dissatisfaction with the programmed material available, had a strong conviction that "programmed learning as a teaching method for languages has no match at the beginner's level."

Consequently, in lieu of the later sections of programmed materials, which were still not forthcoming, one of the Spanish instructors (a native speaker) made 34 tapes based on the approach of the _Spanish A_, which were used along with conventional classes throughout 1963-64, and will be used again in 1964-65.

Experimental Material in German

In comparing the end-of-year performance of the students who started the year with programmed learning it seemed desirable to include in the comparison the performance of students of German who were taught by the use of materials that had been specially prepared and were in some ways analogous to the programmed learning materials. They were prepared as a private venture by Klaus A. Mueller, the Coordinator for this program, F. Alan DuVal of Cornell College and Herbert Wiese of Coe College. One of the authors, Klaus A. Mueller, characterized their approach as:

> . . . a new aspect in learning a language by the audio-lingual method. The objective of this course is the development of all four language skills to the highest possible degree. It is designed for a two-year sequence of language instruction in which the first year is devoted to developing mastery of high-frequency structures, within limited lexical range. The less common and more involved patterns are treated in the second year of the program. The first year text contains reading material carefully selected for simplicity of vocabulary and

structure. Original, but unadapted literary works, poems, short stories and anecdotes are included which meet these standards. More extensive and more challenging readings constitute the basis of the second year program.

The material for the first year is organized into thirty units designed to provide material for a typical week of college level, audio-lingual instruction. Each new structure is presented in the context of old, i.e., previously learned vocabulary. New vocabulary is presented later in each unit in context of previously learned structures. The student is not expected to learn a new structure when it is orally introduced, but to become aware of it and to gain insight inductively into its formation and use. The subsequent drills, exercises, dialogues, and prose passages are designed to implement the active learning and mastery of the new forms and vocabulary.

The designation "Experimental Materials" is used for this approach in the following analysis.

Statistical Analysis

Within college comparison of year-end performance on the MLA Cooperative Language Tests by experimental programmed learning group with a class taught by conventional methods was possible at one college for German and two colleges for Spanish. Similar comparisons for the German experimental materials were feasible at two colleges. The results are shown in Table 27 and 28. In German the only statistically significant difference in scores in the pro-

TABLE 27

AVERAGE ACHIEVEMENT SCORES AND COURSE GRADES BY LANGUAGE AND COLLEGE COMPARISON BETWEEN PROGRAMMED LEARNING AND CONVENTIONAL MATERIALS

1962 - 1963

Language and Treatment	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
German										
Programmed Learning (College II)	19	58	19	35	20	30	20	78	20	3.80
Conventiontal	13	50	13	22	10	26	11	78	15	3.33
Significance		NS		.01		NS		NS		NS
Spanish										
Programmed Learning (College I)	10	46	12	17	12	19	12	54	12	3.42
Conventional	25	50	27	23	29	26	29	75	30	3.93
Significance		NS		.01		.01		.01		NS
Programmed Learning (College II)	23	57	24	23	24	22	24	59	24	3.12
Conventional	23	60	24	24	24	28	24	71	24	3.20
Significance		NS		NS		.05		NS		NS

TABLE 28

AVERAGE ACHIEVEMENT SCORES AND COURSE GRADES BY LANGUAGE AND COLLEGE COMPARISON BETWEEN EXPERIMENTAL AND CONVENTIONAL MATERIALS

1962 - 1963

Language and Treatment	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
German										
Experimental (College III)	20	72	21	28	21	31	21	86	22	3.77
Conventional	15	66	19	33	19	32	19	83	21	3.76
Significance		NS		.05		NS		NS		NS
Experimental (College IV)	23	43	26	29	27	28	27	68	27	3.18
Conventional	40	43	60	22	63	23	63	60	63	3.39
Significance		NS		.01		.01		.05		NS

grammed learning comparison was the higher mean score of the experimental group on the listening comprehension test. In Spanish the three significant differences at College I and the single significant difference at College II all favor the more conventional approach.

The comparison involving the experimental materials resulted in one statistically significant comparison in favor of the conventional approach (listening comprehension) at College III and three significant differences -- listening comprehension, reading, and writing -- in favor of the experimental materials at College IV.

The above comparisons are the only ones possible unless data from different colleges are merged. Such a procedure, of course, introduces many other possible sources of the differences in addition to the two being compared. A comparison of the combined test scores of the programmed learning groups with the mean scores of all of the students in the experimental materials groups involves a comparison of results at Grinnell, Monmouth and St. Olaf with results at Coe and Cornell. Such a comparison is made in Table 29.

TABLE 29

AVERAGE ACHIEVEMENT SCORES AND COURSE GRADES IN GERMAN FOR ALL COLLEGES PARTICIPATING COMPARISON BETWEEN PROGRAMMED LEARNING AND EXPERIMENTAL MATERIALS

1962 - 1963

	Speaking		Listening		Reading		Writing		Grades	
	No.	Mean	No.	Mean	No.	Mean	No.	Mean	No.	Mean
Programmed Learning	32	66	34	29	35	31	34	85	36	3.77
Experimental	42	50	45	31	47	29	47	72	47	3.45
Significance		.01		NS		NS		.01		NS

The programmed learning groups performed significantly better than the experimental materials groups on both the speaking and writing tests, but not on the other measures of performance.

The briefness of the experience with programmed materials makes evaluation difficult. Special tests were given to all of the students in this experiment when the programmed materials were completed. However, there was no valid way of comparing their performance with that of students following other methods. The vocabulary and structural materials covered were less extensive than in the regular courses, but the audio-lingual skills were developed to a more nearly automatic level. At the end of the year these students, along with those participating in other experiments, were given the MLA Cooperative Language Tests.

Evaluation by Teachers

Because the experience with the programmed materials was unfortunately truncated, and the year-end tests a rather dubious measure of either the usefulness of the programs used or the potentialities of the method, it seems desirable to base our conclusions primarily on an analysis of the subjective reactions of the faculty members and students. When the programmed materials had been completed, special questionnaires (shown in Appendix 3) were administered to the students and one faculty member for each experimental group. These questionnaires were filled out by 74 Spanish students, 54 German students and 8 faculty members.

In considering the reactions of faculty members to the programmed learning experiment it is well to keep in mind that participation in this program was voluntary on the part of faculty members and consequently those who took part in these experiments were favorably disposed toward programmed learning, or not adversely disposed toward it, at the outset. Nevertheless, their primary concern was to provide effective learning experiences to their students and it would be unfair to assume that their initial biases unduly influenced their judgments with respect to outcomes.

The eight faculty members were asked to indicate on a rating scale the degree of mastery that they felt that the students with the programmed learning materials achieved as compared with other beginning classes they had taught after a similar length of time. Their responses are summarized in Table 30. The favorable judgment of the programmed learning approach revealed by Table 30 is quite remarkable. It is interesting that all of the instructors felt that these students were superior to other students in the audio-skills of pronounciation and ability to reply quickly and accurately. They were also unanimous in their judgment that they were superior in their control of vocabulary. There was only one dissenter with respect to their superiority in the use of adjectives, the use of possesives, their control of basic structures and their enthusiasm for language learning. Three of the instructors felt that these students were no better than others in the understanding of the instructors'use of the language and in degree of interest. Only one instructor felt that they were inferior to conventionally instructed students in their degree of interest in the language and enthusiasm for

TABLE 30

DISTRIBUTION OF RESPONSES TO THE COMPARISON QUESTIONS BY AREA OF MASTERY AND RATING

	RATING				
Area of Mastery	Much Better	Better	Same	Less Good	Much Less Good
Pronunciation	5	3	0	0	0
Control of Vocabulary	3	5	0	0	0
Correct use of verb forms	1	4	3	0	0
Use of adjectives	3	4	1	0	0
Use of possessives	1	6	1	0	0
Control of basic structures	2	5	1	0	0
Understand *your* use of language	2	3	3	0	0
Reply quickly	3	5	0	0	0
Reply accurately	1	7	0	0	0
Degree of interest	2	2	3	1	0
Enthusiasm for learning the language	2	4	1	1	0
TOTAL RESPONSES	25	48	13	2	0

learning it.

An interesting question with respect to programmed materials is their suitability for students of varying degrees of aptitude for language learning. The instructors' responses to this question are set forth in Table 31. They reveal that all but one of the faculty members felt these materials were superior to conventional materials for the student of low language aptitude. Exactly half of the instructors felt that they were more effective for students of high aptitude but four felt that they were no more effective than other materials, and one felt that they were less effective for students of high ability. The following comments with respect to student mastery give some sense of the enthusiasm of the individual instructors:

> "I think that the programmed learning of Spanish as a teaching technique is wonderful. There is an evident fast mastery of the pronunciation and the basic structures of the language by the student."
>
> "Half of the examination was based exclusively on the programmed materials, and I feel that mastery was extremely high, averaging, roughly estimated, over 90 percent. Also on the rest of the course as I gave it, although we had taken a considerable jump in the rate of introduction of new materials and oral work, the results were well above average for a 'normal'

TABLE 31

DISTRIBUTION OF COMPARISONS OF PROGRAMMED VERSUS NON-PROGRAMMED MATERIALS BY EFFECTIVENESS AND STUDENT APTITUDE

	Effectiveness		
Aptitude	Greater	Same	Less
High	4	3	1
Medium	6	2	0
Low	7	1	0

course. Hence on the plus side of the ledger I would score these points: 1) a solid foundation has been laid for successful further work in the language 2) the students have this feeling too, and morale is high; they feel confident 3) the teacher has been relieved of about half the time he would have spent in the conventional course."

"The results have been unbelievable. . . . Motivation ran very high and students were generally disappointed when the course ended. ... Since the 'traditional' classroom situation has begun several students are having difficulty. ... It is interesting to note that the students are continuing to use the lab faithfully despite having to use less 'interesting' materials. ... I might summarize by saying that I am mildly fanatic about the possibilities for P.L."

"I was amazed at the perfection with which the students mastered their material. ... Although I am sure that the students have not been exposed to as much vocabulary as in a normal text in ten weeks time, I am more than willing to settle for their knowing what they have had and knowing it perfectly and actively (without any groping). . ..I am so enthusiastic about this programmed learning that, if costs are not prohibitive, I want to teach all sections of beginning Spanish this way in the fall and I will use all units that come along: B, C, D, etc. If there is a French program, we would like to use it, too, if it is as good as Spanish 'A'. ... If and when Italian comes along in programmed form, this would, if two years or about that much were available, permit us to offer enough Italian so that we could send a group of students to Italy in connection with our world outlook program."

In summary, it may be said that seven of the eight instructors were quite enthusiastic about the progress made by the students. The eighth instructor was not disillusioned with programmed learning but felt that the program used could have been greatly improved for college students. Unfortunately the questionnaire to the faculty members did not ask for an appraisal of the quality of the program. A number of the teachers commented on this and all of those commenting had suggestions for improvements. The programs were designed for students from junior high school to college and may have been too slow and repetitious at least for the superior students at the college level. It is possible that the instructors' judgments with respect to the superiority of programmed learning for low ability as compared with high ability students may have reflected the characteristics of the particular programs used rather than programmed learning as such.

Five of the eight instructors felt that the students using the programmed materials spent more time than their conventional counterparts in study. One instructor felt that there was no difference in the amount of time spent. One felt that the programmed learning students spent less time and one felt that they spent much less time than the other students in the beginning course.

So far we have been concerned with the instructor's appraisal of the students' experience. We now turn to the instructors' views of what programmed learning means to the teacher. All eight of the respondants indicated that they liked teaching with programmed materials. They split equally between the categories of "very much" and "much" in characterizing the degree of their liking for this approach.

The instructors were asked to give an estimate of the number of students they could handle with programmed materials with the same amount of work as a class of 25 students taught in a conventional manner. It is not surprising that their responses are closely related to the procedures they followed with the programmed materials. One instructor estimated that he could handle no more students with programmed materials than with conventional materials. He estimated that if he had two sections of 25 students each studying with programmed materials he would need to spend about 30 hours a week on their instruction. He would spend eight hours of preparation and test correction. He would meet them in class 2 hours per week and he would spend 20 hours per week administering oral tests.

One instructor felt that he could handle 35 students with the same work as 25. He indicated that the amount of work depends upon student motivation, enthusiastic students requiring less attention. He would have the students scheduled to meet with the instructor once or twice a week to check progress.

Two faculty members felt they could handle 50 students and one 60 students, with the same amount of work as required by 25 students under conventional methods. The two who set their estimates at 50 did not have class meetings with the programmed material students, but had brief conferences averaging once a week to check progress. One of them required students to pass quizzes before advancing to new materials.

Two faculty members felt that five times as many students could be handled by an instructor with programmed materials. One of these met his class once a week and spent the hour doing oral work and administering a twelve minute quiz. He felt that five times as many students could be handled without additional burden only if the students remained somewhat together. If the fast students increased their lead continuously it would become increasingly more time consuming to use materials from several units simultaneously. The other person who estimated that five times as many students could be handled met his programmed materials students twice a month for administrative reasons only. A mid-program examination and a final oral examination were given for grading purposes. He estimated that oral examinations could be administered to a class of 25 to 30 students in about 3 hours. Aside from this he felt that there was little to do with students using programmed materials.

One faculty member felt that an instructor could supervise an unlimited number of students working with programmed materials. The number that could be accommodated would depend only on the physical facilities available and the amount of advanced student assistance available for correcting the weekly tests he would give to the programmed learning students.

Evaluation by Students

It is clear that the faculty members involved in this experiment rated the programmed learning approach very high in terms of student learning outcomes and the burden on the instructor. But how did the students react to the experience? It might be expected that student reaction would be related to their basic motivations for language study.

Student reactions to the programmed materials were on the whole very favorable. This is indicated by their responses to the question how they felt about the programmed learning course. Their answers are tabulated in Table 32. The responses of students to the two language programs were very similar. Over three-quarters of the students in both cases indicated that they either liked the program quite well or liked it very much. In both cases

TABLE 32

DISTRIBUTION OF STUDENT RESPONSES BY NUMBER AND PERCENTAGE TO PROGRAMMED LEARNING

Response	LANGUAGE			
	German		Spanish	
	No.	Per cent	No.	Per cent
Liked very much	14	25.9	20	27.0
Liked quite well	27	50.0	36	48.6
No strong feelings	5	9.3	9	12.2
Not liked very much	6	11.1	7	9.5
Very much dissatisfied	2	3.7	2	2.7
TOTAL	54	100.0	74	100.0

there were two students who were very much dissatisfied with these materials and about 20 percent of the students either had no strongfeelings or did not like the programs very much.

The students were asked two further questions that were much more specific with respect to their attitudes toward the programmed materials. They were asked which they preferred, programmed learning or conventional instruction. They were also asked if further programmed instructional material were available whether they would prefer to continue with it or turn to a more traditional type of course. The responses to these two questions are tabulated in Table 33.

The expressed preferences for programmed learning as compared with conventional instructional materials conform closely to the responses to the more general question as to how well they liked programmed learning. 49 of the 65 Spanish students, or 75 percent, indicated that they preferred programmed learning. Of the German students 32, or 80 percent, expressed this same preference. When they were

TABLE 33

COMPARISON OF STUDENT ATTITUDES TOWARD PROGRAMMED LEARNING AS COMPARED WITH MORE CONVENTIONAL INSTITUTIONAL MATERIALS

	PREFERENCE FOR PROGRAMMED LEARNING OR CONVENTIONAL INSTRUCTION					
	German			Spanish		
	Programmed Learning	Conventional Instruction	TOTAL	Programmed Learning	Conventional Instruction	TOTAL
Preference for further study						
Programmed learning	20	1	21	47	2	49
Traditional course	12	7	19	2	14	16
TOTALS	32	8	40	49	16	65

asked about their preferences for further study of the language, however, a surprising difference developed as between the students of the two languages. In Spanish 49 of the 65, again 75 percent, preferred to continue with programmed materials. Of the 40 German students, however, only 21 or slightly more than half preferred to continue with the programmed materials. The surprising aspect of the responses by German students was that twelve of the 32 who preferred programmed learning to conventional instruction nevertheless wanted no more programmed learning. The reasons for the somewhat inconsistent attitudes of these twelve students and the contrast between the German and Spanish students are not apparent. The reasoning of the German students ran that the "naturalness," "security", "immediate reinforcement," etc., of the programmed materials were desirable for learning basic skills, but the method would become boring if continuedtoo long, and a student-teacher relationship was needed for explanations as the subject became more complicated.

The students who preferred programmed learning over more conventional procedures were asked to give reasons for their preferences. Table 34 shows the categories into which the student responses fell. For both the Spanish and German students the most important things were that they felt they got a better understanding of the language, retained it better, learned more, and found it was a more natural way of learning a foreign language. The other significant type of response reflected the students' interest in being able to speak the language. Of less concern to the students were the facts that it permitted one to work at his own pace, that the programmed materials were easier, or more interesting and enjoyable. Other responses were too infrequent to merit mention.

Since the students' chief concern, quite properly, was in how well they were learning the language, it is interesting to look at their answers to questions dealing directly with this question. They were asked, for example, whether they felt they were learning the language well. The responses of the German students were 48 yes and 6 no. The responses of the Spanish students were 62 yes and 11 no. They

2

TABLE 34

DISTRIBUTION OF RESPONSES OF STUDENTS PREFERRING PROGRAMMED LEARNING BY CATEGORY AND LANGUAGE

Category	Language German	 Spanish
1. Easier	4	3
2. Better understanding, retention; learned more; more natural	12	18
3. Work at own pace	4	8
4. More interesting, enjoyable	1	5
5. Emphasis on pronunciation, speaking	6	14
6. Freedom of scheduling	2	1
7. Painless	0	3
8. Better for beginning course	1	0
9. Eliminates possibility of poor teacher	1	0

were asked further to rate their own mastery of the different aspects of the language on a five-point rating scale running from 0 "very poorly" to 4 "very well". The responses of the German students to this question are shown by college in Table 35. The one aspect of the course which the students consistently rated high was pronunciation. Ratings on the other aspects of programmed learning were not so uniform and generally range between "well" and "very well". The responses of the Spanish students, shown in Table 36 are very similar to those of the German students. The responses for all colleges combined indicate that the students felt they were doing best in pronunciation and next in vocabulary, with some variation in response on the other aspects about which inquiry was made. No average overall rating was less than half way between "well" and "very well".

It is clear that the students have a sense of accomplishment with the programmed materials. There still remains the question of the efficiency of the learning process. Did the students learn more, less, or the same amount with a

TABLE 35

MEAN RATINGS OF PROGRAMMED LEARNING BY GERMAN STUDENTS BY COLLEGE AND ITEM*

Item	Monmouth	St.Olaf	Grinnell	All Colleges
Pronunciation	3.1	3.2	3.3	3.2
Vocabulary	2.4	2.8	3.5	3.0
Verbs	2.5	2.6	3.2	2.7
Adjectives	2.2	2.7	2.5	2.5
Conversation	2.2	2.6	2.8	2.6

* The rating scale was: Very poorly—0; Poorly—1; Fairly well—2; Well—3; Very well—4.

TABLE 36

MEAN RATINGS OF PROGRAMMED LEARNING BY SPANISH STUDENTS BY COLLEGE AND ITEM [1]

Item	Monmouth[2]	St.Olaf	Beloit	Cornell	Knox	All Colleges
Pronunciation	3.0	3.4	3.6	3.3	3.3	3.4
Vocabulary	4.0	2.9	3.6	2.6	3.0	2.9
Verbs	2.0	2.3	2.6	2.7	2.7	2.5
Adjectives	2.0	2.5	3.4	2.9	2.9	2.8
Conversation	3.0	2.4	3.1	2.3	3.0	2.5

[1] The rating scale was: Very poorly—0; Poorly—1; Fairly well—2; Well—3; Very well—4.

[2] Based on only one student.

given expenditure of effort? Of the 40 German students responding 24 said they learned more, four the same, and 12 less, than they would learn with the same expenditure of effort with more conventional materials. 58 Spanish students responded, of which 38 felt they learned more, 11 the same and 9 less than they could have learned with the expenditure of the same amount of effort with other materials. Here again we observe some discrepancy between the German and the Spanish responses. The more versus less ratio for German was 2 to 1 while in Spanish it was more than 4 to 1. A different, but somewhat related question was whether the students found the programmed learning materials easier, harder or of the same difficulty as more conventional materials. Here the responses were more nearly unanimous. 34 of the 40 German students responding indicated that the programmed learning materials were easier, two thought they were harder and four could detect no difference. Of 71 Spanish students responding 68 said they were easier, one said they were harder, and two felt that there was no difference.

Student estimates of the number of hours spent on the programmed material and the number of weeks spent are shown in Tables 37 and 38 respectively. The accuracy of the estimates made by the students undoubtedly varies considerably. In some cases very good records were kept. In other cases the estimates were based simply on recollection. One wonders, for example, whether the actual difference in time spent by students on the same program at different colleges was as great as the means and medians in Table 38 would seem to indicate. Even allowing for individual underestimations and overestimations, it seems clear that there is a considerable range in the amount of time that it takes students to go through a language program. The reported range for German was from 50 hours to 150 hours and for Spanish from 30 hours to 110 hours. Using the traditional formula of one hour of credit for each three hours per week spent on a course, it appears that the German program was approximately equivalent to a three credit course for ten weeks and that the Spanish course was the equivalent of a three credit course for eight weeks. It also appears that, if the students with the lower time estimates are to be

TABLE 37

FREQUENCY, MEAN, MEDIAN, AND RANGE OF NUMBER OF HOURS REQUIRED TO COMPLETE PROGRAMMED MATERIALS BY COLLEGE AND LANGUAGE

	German				Spanish			
College	No.	Mean	Median	Range	No.	Mean	Median	Range
Monmouth	8	73.8	72.5	55- 95	1	75.0	75.0	—
St. Olaf	23	88.5	85.0	50-135	27	71.0	70.0	35-100
Grinnell	14	99.9	100.0	65-150	—	—	—	—
Beloit	—	—	—	—	13	68.5	70.0	30-110
Cornell	—	—	—	—	17	80.8	80.0	48-100
Knox	—	—	—	—	5	56.6	62.0	41- 70
All Colleges	45	89.4	85	50-150	63	72.0	70.0	30-110

TABLE 38

FREQUENCY, MEAN, MEDIAN, AND RANGE OF NUMBER OF WEEKS REQUIRED TO COMPLETE PROGRAMMED MATERIALS BY COLLEGE AND LANGUAGE

	German				Spanish			
College	No.	Mean	Median	Range	No.	Mean	Median	Range
Monmouth	10	9.0	10	7-10	1	9.0	9	—
St. Olaf	24	10.0	10	9-11	28	9.9	10	8-12
Grinnell	19	9.9	10	8-11	—	—	—	—
Beloit	—	—	—	—	13	10.1	10	9-12
Cornell	—	—	—	—	21	12.1	12	9-18
Knox	—	—	—	—	7	14.6	15	10-18
All Colleges	53	9.8	10	7-11	70	11.1	10	8-18

believed, some of them in German might have been expected to put in up to something less than twice as much time and in Spanish up to three times as much time as they did on the course. If these figures are to be accepted, there is clear need for arrangements that will permit the quicker, high aptitude students to cover more ground in a given period of time than they were required to with these programs.

In general the programmed materials were expected to take about ten weeks of time and in many cases schedules were set up that worked toward this end. At Knox a good deal of supplementary material was provided to the students and the use of the programmed materials spaced out over a longer period. The statistics presented in Table 38 based on the student estimates of the number of weeks required to complete the course are probably considerably more accurate than the estimates of the number of hours shown in Table 37. It may be seen that the ten week projection was reasonably adhered to. In this case, also, the range is much less. Apparently the fastest students paced themselves week by week so as not to complete the course too much ahead of the other members of the class.

The fact that the students working with the programmed learning materials were not overworked is indicated by their responses to the following question: "In your period of study with the programmed learning course did you put in more or less work as compared with previous language courses during a comparable period?" Fifteen of the German students said they spent more time, eleven the same and thirteen less time than with previous language courses. The results for the Spanish students were ten more, twenty the same and twenty-nine less time on the programmed materials. This is further evidence that, in terms of time required, the Spanish program was less demanding than the German program and also that a good many students might have properly been asked to do more work than they did.

The overwhelming judgment that the programmed materials were easier than conventional material naturally leads to the question why the students felt this way. Their responses are categorized and tabulated in Table 39. The two

TABLE 39

DISTRIBUTION OF RESPONSES OF STUDENTS AS TO WHY PROGRAMMED LEARNING MATERIALS WERE EASIER THAN CONVENTIONAL MATERIALS, BY CATEGORY AND LANGUAGE

Category	Responses* German	Responses* Spanish
1. More interesting	0	3
2. Repetition	13	12
3. Less to memorize, cover, or learn	0	9
4. Easier to concentrate	0	2
5. Learned faster	0	1
6. Clearer, more natural	12	15
7. Learn at own rate	1	3
8. Smaller bits presented	0	4
9. Less wasted time	0	2
10. Very elementary	0	2
11. Easier to keep up	2	1
12. Slower	3	4
13. Hearing the language	1	1
14. Less study required	4	1
15. Not as subject to structure	0	1
16. No fear of quiz	1	0
17. Easy comparison with French and Latin mechanics	0	1

* Students might respond with more than one reason. The numbers shown are, therefore, number of responses, not number of students.

characteristics of the programmed materials that stand out in this presentation are the amount of repetition in the programmed materials and the students' feeling that they were clearer and that learning with them was more natural than with other materials. An interesting contrast between the Spanish and German students is found in the fact that nine Spanish students felt that there was less to memorize, cover, or learn in the Spanish programmed materials, but no German student made this comment.

The students were asked two questions with respect to the materials and procedures followed. In response to the question whether they found the quality of the tapes to be good or bad all 54 of the German students responding replied good. The Spanish students response was 67 good and 6 poor. There was apparently little dissatisfaction with the technical qualities of the materials that the students worked with.

The general approach of the programmed learning materials to grammar instruction was "inductive" and there was a minimum of explanation of grammar in conventional terms. With respect to this the students were asked the following question: " Do you think that the absence of detailed grammar explanations in the programmed learning course has been an asset or a drawback in your gaining of initial competence of the language?" The responses to this question are shown in Table 40. There is some contrast here between the German students and the Spanish students. Half of the German students felt that the absence of the grammatical explanations was a drawback and approximately a quarter felt that this was an asset. Less than a third of the Spanish students felt that this was a drawback and approximately a fourth thought that this was an asset. The neutral categories of "neither" or "don't know" were relatively more common among Spanish students than among those who studied the programmed materials in German.

To a more fundamental question on the usefulness of the group meetings of the class the student responses were much more varied. The responses of the students are shown in

TABLE 40

DISTRIBUTION OF RESPONSES TO QUESTION ON THE ABSENCE OF EXPLANATIONS OF GRAMMAR BY: RESPONSE AND LANGUAGE

Response	Language German	 Spanish
Asset	9	16
Drawback	20	18
Neither	6	9
Don't know	5	17
TOTAL	40	60

Table 41. There is an interesting contrast between the responses of the German and Spanish students. 39, or 72 percent, of the German students felt that the group meetings were either useful or highly useful. Among the Spanish students only 38 of 74, or slightly more than 51 percent gave this favorable response. 36 felt that the group meetings were either not useful or were a waste of time.

The students who felt that the group meetings were either useful or highly useful were asked to comment on this point. The responses given are categorized in Table 42. They indicate that the most generally served purpose, in the students' estimation, was the clearing up of questions with respect to language study. The coordination and comparison of student progress and assistance with conversation and pronunciation were also indicated with some frequency. Six of the German students also felt that the testing aspect was useful. Other types of comments were rather infrequently made.

TABLE 41

DISTRIBUTION OF STUDENT FEELING ABOUT GROUP MEETINGS;
BY LANGUAGE AND RESPONSE

	Language	
Response	German	Spanish
Highly useful	9	6
Useful	30	32
Not useful	12	26
A waste	3	10

TABLE 42

DISTRIBUTION OF USEFUL FEATURES OF THE GROUP MEETINGS BY CATEGORY AND LANGUAGE

Feature	German	Spanish
1. Clear up questions relative to language study	15	22
2. Administrative questions answered	1	2
3. Coordinate and compare student progress	10	5
4. Conversation and pronunciation	8	6
5. Application and practice	4	2
6. Motivational effect	0	2
7. Testing	6	2
8. Change in atmosphere	1	0

Conclusions

For the faculty members who participated in them the experiments with programmed learning of foreign languages was at the same time a most rewarding and frustrating experience. It was rewarding because of the potentialities that were revealed, and frustrating because the slow rate of production of programs makes it clear that it will be a long time before the potentialities can be significantly realized.

The production of programmed materials for language learning is a difficult art, and there is considerable room for improvement in the materials that were used in these experiments. The results obtained, however, make it quite clear that students using these materials do learn relatively well and quickly and that substantial economies in instructorial time and effort can be realized. There is also the prospect that programmed learning will make much more feasible greater individualization of instruction, relating each student's learning experiences more directly to his linguistic ability and state of mastery of the language.

Programs are needed covering the full range of materials involved in foreign language study. There should be programs that are designed for different educational levels, and at each level there should be faster programs for the linguistically talented and slower programs for those to whom language learning is not so easy.

With such a set of programs a number of potentially significant experiments could be carried out. Procedures should be developed for determining the appropriate pace for the individual student. Explorations should be carried out of the possibility of grouping students by stage of mastery into formal or informal groups for language experiences on a more personal level than the programmed materials permit. Since the degree of mastery of the materials covered seems very high for students in programmed learning the conventional procedure for organizing courses so that students achieve varying degrees of mastery of a large body of materials might be replaced by one in which all students achieve nearly complete mastery of the materials but cover varying amounts of material.

Programmed learning suggests a most promising way of dealing with the placement of students with previous language training. A combination of tests and review programs should make it possible quickly to establish appropriate placement both in level and pace of the materials to be studied. The wasteful -- or even harmful -- effects of being placed with a group of lesser language competence could be avoided as well as the frustrating experience of competing with students more advanced and/or linguistically talented than one's self.

It may be conceded that program construction is a difficult art, calling for many talents. One may still wonder whether the mystique has not been overdone. To a much greater extent than with other types of learning materials programs are empirically derived and can be readily tested. Their full potentialities will not be realized until large num-

bers of language teachers have been inducted into the mysteries of program construction and have become involved in the production of programmed materials for both local and more widespread use.

VIII. PACE AND MASTERY

If dependence were to be placed entirely on programmed learning of a language the result would be that each student would proceed at his own pace, but eventually all students would achieve practically the same degree of mastery of the language. In class instruction, on the other hand, all students proceed at approximately the same pace, measured by vocabulary and structural items introduced, but with a wide range in mastery. Where, as in most of the ACM colleges that experimented with programmed learning, there were class meetings and frequent quizzes, there might be considerable variation in the time spent in study each week, but the students presumably proceeded at approximately the same pace and with roughly the same degree of mastery.

The problems that arise from a considerable range in mastery in class instruction are so significant that there was a good deal of interest shown in this problem in the first Summer Workshop, in 1960. The question was raised whether results might be improved if, in class instruction, a certain minimum level of mastery of materials already presented was required of the students before proceeding to new words and grammatical constructions. This idea was related to procedures in several special language centers such as the Army Language School. There very small groups -- usually of three persons each -- over-learn material, even to the point of perfect rote memory of the conversation units; and every individual is expected to have 100 percent mastery of the material covered. Obviously such a plan would require major alteration both in philosophy and objectives in order to fit into a college teaching situation.

Nonetheless it was agreed that an experimental design should be studied, with the stated purpose of determining "whether a high degree of demonstrated mastery of each lesson unit before a new unit is introduced will achieve better results in the mastery of the four language skills at various stages and at the conclusion of the course."

Three major problems soon made themselves apparent.

One was the balance between the optimum progress of the individual student (in line with the philosophy of the individualization of instruction) on the one hand, as against the practical necessities of handling students in relatively homogenous groups. If the members of a class started out more or less on a par, and then were tested for a specified level of mastery, the class would have to be sub-divided if any recognition was to be made that one group was prepared to proceed while another needed more time. This process might well end up with a scheduled class being fragmented into several sub-levels, one or more of which might possibly include only one student.

Such a situation is clearly impractical in terms of administrative complexity, classroom space, teacher availability, efficiency of teacher activity (de facto tutorial systems are wasteful and expensive in this respect), evaluation standards, relative credit, the difficulty of devising numerous different tests to assess the same level of proficiency on the same materials, etc. Yet allowing the individual to proceed at his own rate of mastery is the reasonable application of such an approach.

A second problem was the degree of mastery to be accepted as the basis for proceeding to new units of material. Many examples of programmed learning set 95 percent as the norm, but is this not too high for a skill subject like beginning language? Might not 70 percent, or 80, or 90 (for the "high level" mentality), be preferable?

The third problem was the difficulty of accurately testing mastery. The absolute validity of any testing device is precarious. Even the best tests are to greater or lesser degree biased to favor some instructional techniques over others. Some very important elements in language proficiency are difficult to assess accurately, and others well-nigh impossible. Even by-passing the last two problems of degree of mastery and flexible but valid testing, the operational difficulties remained a major obstacle.

However, the membership of the 1960 Workshop felt such keen interest in the problem that an experimental design was worked out. The design perforce bowed to the pressure of operational realism, and was applied on the basis that 75 percent of the experimental class must have mastered 80 percent of the material before moving on to the next unit of material. This experiment was tried in only one college, in two sections.

The reasons for calling off the experiment geared to mastery were expressed in the reaction of the language faculty of the college that, if continued, it would "destroy the academic worth of the class", and in the judgment of the experienced teacher in charge that the application of this approach in a group manner was "unrealistic from both the psychological and academic viewpoints." The teacher further reported that "students who were told to repeat study of materials already assigned not only did not do so but did less well upon return to reassigned material in class. Students were quick to discover a means of slowing down the pace and content of the course,and, if given the opportunity, would have brought the course to a complete standstill."

Two implications of this scheme deserve brief mention.

One is the impingement on the characteristics of the programmed learning approach. In the conventional classroom procedure time is the constant, usually in the form of hours of class and laboratory translated into credit-hours, while mastery by the individual student is the variable, as shown in the grade distribution of the individuals making up the class. In the programmed learning approach, this relationship is reversed; time becomes the variable, in that the individual student works at whatever speed is appropriate until he has reached a certain degree of mastery, the amount of allowable error being decided in advance by the programmers and built into the sequence.

The second implication has to do with the capacity of the individual student to master a language. All college grading says, in effect, "Because of constraining circumstances, we will move a student ahead who never achieves more than 70 percent mastery, whether he is capable of learning the material or not."

The plan of moving the entire group ahead when 75 percent have mastered 80 percent of the material contains at least the suggestion that some students just never will attain full mastery. Certainly faculties giving such wholehearted endorsement to the values of language study as to require a certain amount for graduation would not officially have supported such a thesis, and no such cynicism is imputed to the teachers involved. Theirs was a pragmatic compromise, but it raises some questions as to the validity of a language requirement.

The experiment was not revived, and in terms of outcome it was of no importance. Its chief value is as a testimony of the serious interest of the language faculites in the problem of optimum mastery. The later development of an experiment in the use of programmed learning was the best replacement that could be realistically pursued in this area. The various national figures who participated as consultants in the Workshops shared the interest in the problem, but in their turn were not very productive of practical designs. The question remains a challenging and significant one: what is the optimum mastery that should be the basis for moving to new material, and how can such mastery be induced?

IX. ATTITUDES, MOTIVATIONS, AND THINKING IN A NEW LANGUAGE

Psychological Factors in Language Study

The conducting of experimental work on student achievement in language under varying conditions would seem arid and incomplete without some attention to the "humanizing" factors in language learning: What effects do the various laboratory conditions and practices seem to have upon student motivations and attitudes? A teacher relies upon feedback in some form in order to remain effective. He has to know when he has been pressing his students too hard or not hard enough; he has to know when their interest is flagging and when they are eager for explanations or discussion -- when they do best by being left to themselves and when they need expert guidance. The psychological testing program of ACM developed partly as a result of pedagogical instincts that demanded some feedback of this kind on the methods and materials that had been used. Moreover the timeliness of such testing was evident in light of the work under U.S.O.E. Contract 8823 of Professor George A. C. Scherer and his colleagues at the University of Colorado.[1] In many respects the Colorado study could be utilized and compared with the results at the colleges of ACM. Professor Michael Wertheimer of the University of Colorado consulted with the Principal Investigator of the ACM language program and with Dr. William Wiersma, the statistical consultant, and so did Dr. Wallace Lambert, who constructed several of the instruments used in the Colorado study. Although the present investigation involved much less extensive post-testing than was done at Colorado, several of the Colorado instruments were used, either directly or with appropriate modifications.

In the adapted forms of Colorado psychological tests used by ACM, the measures provided 19 variables. These in turn related to three areas of concern: there were measures of the kind and extent of thinking done in the foreign language, measures of motivation, and measures of the degree of cultural orientation. Several interesting instruments used at the University of Colorado were omitted in the ACM program on the advice of Professors Wertheimer and Wiersma. Instruments were omitted that had not proved valuable in the Colorado study. Other considerations were that a large bulk of materials was not required, and that it was desirable not to impose too much on a group of students and teachers that had already been subjected to a good deal of examining and "temperature-taking". Those Colorado instruments that were borrowed with or without changes by ACM were I, V, VI, XII, XIII, and T, in terms of the system of test designation used in the Colorado Study.

---

[1] The report is now available in published form: George A. Scherer and Michael Wertheimer, *A Psycholinguistic Experiment in Foreign Language Teaching*, 1964, McGraw-Hill.

I. A) B)	Lambert's tests of "instrumental" versus "integrative" purpose
V.	"Germanophilia"
V-B.	"Anomie"
VI.	Motivational strength
XI.	Attitudes toward language-centered concepts
XIII.	Semantic differences
T	Measures of the degree of thinking in the language.

The actual instruments used by ACM are shown in

Appendix 4. According to the way in which results were tabulated into 19 measures by ACM, the variables may be classified as follows:

I. Measures of degrees of thinking in the foreign language

Item Number	Measure	Source
1	Sensitivity to semantic differences in the foreign language	Reduction of Colorado Form XIII; total score on 11 semantic judgments
2-7	Self-assessment of ability to think in the foreign language	Duplication of Colorado T - test
(2)	*Need to interpose English translation*	*T-1*
(3)	*English intrusions when speaking foreign language*	*T-2*
(4)	*Foreign language intrusions, when speaking English*	*T-3 and 4*
(5)	*Foreign language in dreams*	*T-5 and 6*
(6)	*Capability to think in foreign language*	*T-7*
(7)	*Total score*	

II. Measures of motivation

Item Number	Measure	Source
8-9	Assessment of "utilitarian" versus non-utilitarian or "integrative" motivation	
(8)	*Score on short Lambert test*	*Colorado IA, "Orientation Index" short form*

Item Number	Measure	Source
(9)	*Score on long Lambert test*	*Colorado IB, "Orientation Index" long form*
10-11	Assessment of strength of motivation	
(10)	*Students' report of hours of language study outside of class*	*Item 1 on Colorado Form VI*
(11)	*Students' total score on items 2-14 of questionnaire on eagerness to learn the language*	*Colorado Form VI, 2-14*

III. Cultural orientation: student sympathy with foreign people, their form of government, and general acceptance of their ethos.

Item Number	Measure	Source
12	Anomie or degree of dissociation from one's own culture	Score on Colorado Form V-B (13-item series, each with a seven-point scale)
13	Phylophilia - fondness for the culture connected with the language studied	Total score on Colorado Form V, "Germanophilia," 1-25
14-19	Attitudes toward language centered concepts	Selected items from Colorado Test XI
(14)	*The people in their own country (as "German people in Germany," e.g.)*	*Form XI, #2*
(15)	*The people (Frenchmen, Germans or Spaniards, whichever appropriate) in this country*	*Form XI, #7*
(16)	*"My language teacher"*	*Form XI, #8*
(17)	*Attitude toward myself speaking English*	*Form XI, #9*
(18)	*Attitude toward myself speaking the foreign language (French, German, or Spanish, whichever is appropriate)*	*Form XI, #10*
(19)	*Difference score; i.e., sum of differences between parts of Items 17 & 18*	

Given these three groups of measures, six kinds of comparisons were explored.

1. Pre-test measures of aptitude and the psychological test scores were compared with post-test measures of success.

2. ACM results were compared with those of the University of Colorado.

3. The differences between colleges in the variables tested were examined.

4. The scores made on the various psychological measures by persons who had not yet studied foreign language at the college level were compared with scores of persons who had completed one year of college language study.

5. The relations between the scores of students who belonged to experimental groups and those in control groups were explored. This was a major aspect of the Colorado investigation.

6. The interrelationships between the different psychometric instruments were investigated.

In Appendix 4 are samples of all instruments used.

Forms were prepared for French, German, and Spanish classes, and administered to about 1,800 students in eight colleges. Two of the eight colleges were omitted from some parts of the analysis because these colleges had participating totals of only 14 students each. As a control measure, some 200 entering freshmen were tested in the early fall of 1963, and their mean scores are used in some analyses as an indication of "pre-college" attitudes, from which the degree of change could thus be at least roughly estimated.

Relations between Psychological Measures and Pre- and Post-tests

Every language teacher naturally seeks to be alert to recognize interrelationships between his students' aptitude for languages and/or mastery of language skills on one hand and certain "psychological" factors on the other hand -- attitudes and motivations of students, their ability to lose themselves, so to speak, in an acquired language, and their cultural orientation. Such relationships are not only difficult to recognize, they are even harder to state in provable form. Indirect approaches to these questions in the recent past may have accomplished more than frontal attack. Such an indirect approach was taken in the language research program of the University of Colorado. Ostensibly the design of experimentation was to measure the relative effectiveness of a "linguistic approach"; but indirectly, there appeared some interesting results as by-products of the effort. The usefulness of these results is not that of well-established and thoroughly tested facts, but that of suggestive clues, corroborated to a degree by ACM studies, as will be indicated below. As suggestions they call attention to further experimental possibilities.

Is there a correlation between one's ability to think in a language (or, as it was more precisely stated in the Colorado study, the degree of habituation of direct associations in the language) and one's mastery of basic skills? Although most of the correlations between "Habituated Direct Association" and aptitude or proficiency scores obtained in the Colorado study were quite low, a correlation coefficient of .40 was obtained between the total scores on the "T" test (the "T" test is the subjective appraisal of one's own degree of "thinking in the language") and the combined scores for proficiency in the four basic skills as these were tested at the University of Colorado. (See Table VI-18, Colorado Report). There was only a very slight relationship between aptitude and subjective appraisal of habituated direct associations (r = .16). The fourth semester correlation with skills was .46 rather than .40 but the correlation with aptitude remained substantially unchanged at .15 (See Table 22, Colorado Report).

One so-called objective measure of habituated direct association was the test of sensitivity to semantic differences. The Colorado "objective" and "subjective" tests were found to have little or no correlation with one another (See Table VI-19, Colorado Report) and it appeared that "sensitivity to semantic differences" is even quite separate from "habituated direct association." Slight evidence exists that sensitivity to the shaded meanings of words has a bearing on the long-range learning of languages (See Table VI-11, Colorado Report, e.g.). The strongest correlation shown was in the fourth semester of the project, when a correlation of .29 appeared between "vocabulary" and the "direction score" in "sensitivity to semantic differences" (Table VI-16, Colorado Report), as evidence which perhaps indicates that semantic orientation may have some slight long-range significance in mastery of language.

What about the relations between "motivation" and the mastery of language and between "cultural orientation" and the mastery of the language? The Colorado study (Table VI-27, Colorado Report) indicates that there were quite low but mostly negative correlations between "instrumental" and personal motives and proficiency in language, and moderate positive correlations between a stated desire for more training and proficiencies in the four basic skills. The correlation between language aptitude and motivation was not studied. With respect to "cultural orientation" -- here defined as the tendency to accept sympathetically the foreign ethos, the government, and the foreign social values -- the Colorado study shows (Table VI-32) generally low correlations, the most clear-cut of which are negative -- thus, a sharp difference between a student's view of himself as a foreign speaker and his view of himself as an English speaker correlated -.26 with total proficiency scores.

Now, the question arises as to whether all these results reported to date were corroborated by ACM studies. With respect to the relations between "T" test scores (based on answers to questions about the degree to which one thinks habitually in the foreign language) and aptitude and proficiency scores, it was felt in this study that non-parametric procedures should be used for analysis, inasmuch as the numerical values of "T" scores had been arbitrarily assigned, and were not necessarily consistently meaningful in the intervals they expressed.[1] Thus the relationships between "thinking in the language" (as rated subjectively) and "proficiency" and "aptitude" were not described in terms of correlation coefficients in the ACM analysis, and comparisons with the Colorado Report in terms of such coefficients is not possible. But in connection with an objective measure of "thinking in the language," the test of sensitivity to semantic differences, ACM can offer a sample study of the control group of German students in the programmed learning experiment. This group of 73 students was used for comparison with the Colorado study because the latter study was confined to German instruction. The correlation between semantic sensitivity scores and other measures were as follows:

---

[1] Mr. Wiersma notes:
"It was not felt that the parametric assumptions could be met. One of the parametric assumptions is that the variables (measures) involved must have been measured on at least an interval scale. The responses to these eight tests were on rating scales and as such there is no basis for assuming that the possible choices were on an interval scale. An interval scale has the characteristic of the distances between any two numbers on the scale being of known size. The various choices were assigned numerical values but only for tabulation purposes."

	ACM	Colorado Report
SAT-mathematics	-.09	.06
SAT-verbal	-.09	.07
MLAT	-.08	.16
Speaking achievement	.00	.43
Listening achievement	-.11	.37
Reading achievement	-.13	.22
Writing achievement	-.14	.21
Grades--2nd semester	-.10	

There is no verification in ACM studies of the positive correlations between achievement tests and "semantic sensitivity"; the ACM results suggest that there may in fact be a negative correlation.

With respect to the Colorado findings on "instrumental" motivation, the ACM investigations did not involve the application of correlation procedures to the data provided by the Lambert scales. Information can be provided, however, concerning other motivational factors and their relation to achievement. One of the component parts of the Colorado total motivation index was the "hours of study" reported by students as to the amount of time spent in study of the foreign language each week. In the German control group of 73 mentioned earlier, correlations were made between pre-test and post-test scores and the "hours of study" estimates as well as the total score on the Colorado instrument for measuring motivation. The correlations are as follows:

	ACM Hours of Study	ACM Motivation	Colorado Report** Motivation
SAT-Math	-.05	.05	
SAT-Verbal	-.03	.05	
MLAT	.46*	.24*	
Speaking Achievement	.16	.35*	.22
Listening Achievement	.13	.11	.09
Reading Achievement	.07	.19	.15
Writing Achievement	.44*	.23	.26
Grades, second semester	.23	.57*	

---

* Significant at .05 or better.

** Table VI-27, Colorado Report.

The significant figures tend to corroborate the Colorado results, and supplement them to the extent of suggesting a correlation between aptitude (as measured by MLAT) and motivation (as measured by "hours of weekly study"). The coefficient of correlation is .46. It also appears that "hours of study" relate more clearly to achievement in writing than to any other of the four skills and that strong general motivation is most obviously related to speaking achievement and to grades.

Relations between Psychological Measures and College Differences

The ACM study included an appraisal of differences among individual colleges; this was an attempt to answer the question of whether there are significant differences in psychological factors, on a college-by-college basis.

The answer to this question is that there are many significant differences. Table 43 shows the differences for eleven of the psychological factors (those that can be handled on a simple numerical basis) in terms of eight colleges and the pre-college sampling. The mean scores are reported.

TABLE 43

Mean Scores on Psychological Tests by Colleges

COLLEGE	I	II	III	IV	V	VI	VII	VIII	Pre-Col-lège
Semantic Sensitivity	21	17	17	22	21	26	20	19	19
Hours of Study	7	6	9	7	6	8	10	9	7
Motivation	14	13	14	14	14	14	15	15	15
Anomie	30	28	29	30	27	28	27	32	29
Phylophilia	81	79	84	80	83	77	80	83	83
Inventory:									
Foreign People in their country	87	74	87	84	93	83	87	82	87
Foreign People in this Country	85	79	84	80	89	86	80	79	82
My FL Teacher	99	88	95	95	99	102	97	92	99
Me Speaking English	93	85	89	87	92	87	89	89	89
Me Speaking FL	76	68	78	73	77	72	80	71	78
Diff. of last two	27	28	17	25	25	28	22	26	22
RANGE IN NUMBERS	280-303	14-15	156-162	232-242	607-641	12-14	288-302	117-123	93-209

Virtually all categories produce significant differences. A calculation of the significance of the difference between the highest and lowest colleges on the anomie scores, for example, showed the significance to be at the .001 level or better. Although the differences in scores are significant, there is no clear-cut pattern in which one college stands out above all the others in scores on all the tests. All but one of the colleges has a top score on at least one item. One college, number II, had only 14 to 15 students taking these tests, and they made, on the whole, rather low scores.

Relations between Psychological Measures and Experimental Studies

A series of correlations were run to see whether there was evidence that the experiments in withholding orthography in laboratory monitoring, in type of laboratory or in programmed learning had discernable effects on the interrelations between the various measures. The detailed tables of correlation coefficients are shown in Appendix 4 (Tables 1-7). Nineteen variables are shown:

1. Semantic sensitivity
2. Motivation: hours studied
3. Motivation: total score
4. Attitude toward native culture: anomie
5. Attitude toward foreign culture: phylophilia
6. Attitudes: Foreigners in their country
7. Attitudes: Foreigners in this country
8. Attitudes: "My foreign language teacher"
9. Attitudes: "Me speaking English"
10. Attitudes: "Me speaking foreign language"
11. Attitudes: Differences between #9 and #10
12. Scholastic Aptitude Test: Mathematics
13. Scholastic Aptitude Test: Verbal
14. Modern Language Aptitude Test
15. Speaking Achievement
16. Listening Achievement
17. Reading Achievement
18. Writing Achievement
19. Second semester grades

The correlation coefficients selected for presentation in Appendix 4 show the relations between phylophilia, the total motivation scores, and anomie and the other variables. The following tables are included:

1. Correlations between Phylophilia Scores and Other Variables for the Withholding Orthography Experiment, June 1963
2. Correlations between Total Motivation Scores and Other Variables for the Withholding Orthography Experiment, June 1963
3. Correlations between Anomie Scores and Other Variables for the Withholding Orthography Experiment, June 1963
4. Correlations between Phylophilia Scores and Other Variables for the Monitoring Experiment, June 1963
5. Correlations between Anomie Scores and other Variables for the Monitoring Experiment, June 1963
6. Correlations between Total Motivation Scores and Other Variables for the Monitoring Experiment, June 1963
7. Correlations between Phylophilia Anomie, and Total Motivation Scores and Other Variables for the Type of Laboratory Experiment, June 1963

In these correlation analyses there is very little evidence of relation between phylophilia or anomie and any of the achievement tests. Some fairly substantial coefficients were recorded between these two attitude measures and some of the pre-test scores, but no consistent pattern was discernable. In the groups involved in the experiment in withholding orthography there were some not-unexpected positive coefficients between phylophilia and attitudes to foreigners both in this country and in their own country. These relations were stronger for the experimental than for the control groups. This relation, however, did not hold for the monitoring experiment.

It is interesting to note that the correlations between phylophilia and anomie are generally positive. Apparently the attraction of a foreign culture has some tendency to be associated with dissociation from one's native culture.

It would be expected that high positive correlations should be found between the total motivation index and achievement scores. A number of such coefficients appeared, but there was no consistent pattern, and even a few negative correlations.

No regular pattern of differences between experimental and control groups was observed, and consequently no generalization is possible with respect to the results of experimentation on relationships between the nineteen variables listed above.

The control group of 73 German students was used to examine the intercorrelations between the attitude scores. The results are shown in Table 44. The intercorrelations of the attitude measures as indicated in Table 44 were generally positive and quite large. The only two not significantly different from zero were the .21's. Both of these appeared with the last variable which was a difference score. The intercorrelations of the attitude measures as reported in the Colorado study were generally positive but of less magnitude than those of Table 44.

The interrelations between pre-test measures and achievement scores are shown in Table 45. There is an unusually high correlation between the two College Board Scholastic Aptitude Tests (.93) and a very modest but statistically significant correlation between the Verbal SAT and the Modern Language Aptitude Test. Neither of the SAT tests show significant relations with achievement. The MLAT shows significant correlations with listening comprehension and writing, but not with speaking and reading. Except for

Table 44

Correlations Between Attitude Measures; German Control Group; June 1963

Variable	Germans in Germany	Germans in This Country	My German Teacher	Me Speaking English
Germans in this Country	75			
My German Teacher	69	71		
Me Speaking English	79	66	70	
Me Speaking German	52	47	55	72
# 13 above - #4	30	28	21*	21*

* Not significantly different from zero at the 5% level.

Table 45

Intercorrelations of Pre-Test and Achievement Measures;
German Control Group; June 1963

Variable	SAT Math	SAT Verbal	MLAT	Speaking	Listening	Reading	Writing
SAT Verbal	93*						
MLAT	22	28*					
Speaking	-01	-01	05				
Listening	-20	-16	33*	22			
Reading	-09	-01	17	09	46*		
Writing	-23	-19	75*	13	55*	37*	
Course Grades	10	08	17	29*	43*	28*	21

* Significantly different from zero at the .05 level

speaking the achievement measures are significantly intercorrelated.

Although the interrelationships between the various measures used are of interest, more significance attaches to the question whether the instructional process resulted in changes in the performance or attitudes of students, as measured by these instruments. To answer this question mean scores on each of the instruments were computed for three groups, a pre-college group of some 200 freshmen, who were tested just as they were entering college, and for each experiment the experimental and control groups. The numbers of students for whom scores were obtained for each experiment are shown in Tables 8 to 11 inclusive in Appendix 4. The number of scores on the different tests varies from variable to variable because some of the students failed to respond to some of the questions.

A comparison of means between the experimental and control groups for the withholding orthography experiment is given in Table 46. It may be seen that there are only three statistically significant differences. The control

Table 46

Means, F-Ratios, and Frequencies for the Psychometric Tests; Withholding Orthography Experiment: June 1963

	Mean			
	Experimental	Control	F-Ratio	N
Phylophilia	81.5	80.9	—	224
Anomie	27.6	31.5	8.87*	225
Hours Study	6.4	7.9	9.03*	221
Total Motivation	13.3	13.7	1.11	224
Foreigners Foreign Country	83.8	84.6	—	221
Foreigners This Country	81.5	82.7	—	217
My FL Teacher	99.0	96.1	2.03	222
Me Spkg. Eng.	89.1	87.5	—	220
Me Spkg. F.L.	74.2	73.6	—	218
Diff. Score	26.0	26.7	—	211
Seman. Sens.	19.7	22.4	5.76*	217

*Significant at the .05 level

group apparently developed greater anomie -- or dissociation from their native cultural environment. The withholding orthography group put in significantly fewer hours of study per week -- according to their own reports. This raises the interesting question whether the lack of the more conventional study materials in the earliest part of the course established study patterns which persisted through the rest of the course. The fact that the control group scored higher in semantic sensitivity than the experimental group is puzzling.

The results for the experiment involving monitored versus non-monitored laboratories are reported in Table 47 The experimental group performed their laboratory assignments without monitoring by teachers or assistants. They report less time spent in study than did the students with monitored laboratories. For some reason they seem also to have acquired a more positive attitude toward foreigners than did the students in monitored laboratories.

Comparisons are given in Table 48 for the experiment involving two types of laboratory. The students who

Table 47

Means, F-Ratios, and Frequencies for Psychometric Tests;
Monitoring Experiment; June 1963

Test	Mean		F-Ratio	N
	Experimental	Control		
Phylophilia	80.7	82.7	1.66	274
Anomie	29.8	31.9	2.72	274
Hours Study	7.1	9.5	20.22*	265
Total Motivation	13.5	14.9	1.67	271
Foreigners Foreign Country	86.5	81.6	6.70*	267
Foreigners This Country	84.9	77.0	13.69*	267
My FL teacher	96.3	93.8	1.32	268
Me Spkg. Eng.	89.2	89.6	—	265
Me Spkg. F.L.	72.9	70.7	—	262
Diff Score.	28.4	29.4	—	253
Seman. Sens.	21.6	19.6	3.50	268

---

*Significant at the .05 level.

Table 48

Means, F-Ratios, and Frequencies for Psychometric Tests; Laboratory Type Experiment, June 1963

Test	Mean		F-Ratio	N
	Non-recording	Recording		
Phylophilia	86.7	81.1	4.94*	691
Anomie	31.4	28.0	2.68	691
Hours Study	9.8	6.8	21.68*	675
Total Motivation	15.9	14.1	1.82	680
Foreigners Foreign Country	92.9	88.8	1.11	678
Foreigners This Country	89.3	83.8	3.87*	663
My FL Teacher	100.3	99.4	—	677
Me Spkg. Eng.	92.1	92.1	—	670
Me Spkg. F. L.	77.3	77.5	—	659
Diff. Score	24.7	26.1	—	637
Seman. Sens.	19.5	20.8	3.51	662

* Significant at the .05 level.

worked in non-recording laboratories reported more time spent in study and seemed more favorably disposed toward the foreign country and toward foreigners in this country, than were the students who did their laboratory work in the more flexible laboratories. In this, as in the other data given in this section, it is well to remember that the students in the two types of laboratories were studying in different colleges, and inter-college differences other than the experimental differences may contribute significantly to the results.

The outcomes for the experiment involving programmed learning materials in German and Spanish, and special experimental materials in German, are shown in Table 48. Both of the experimental groups reported more time spent in study than did the control group. In total motivation score the experimental materials groups made a significantly higher score than the control group. The differences between the programmed learning students and the other two groups on this motivational score were not significant. On the scale reflecting attitude toward oneself speaking English the control group had the highest score and the programmed learning group the lowest. The import of this finding is not apparent.

The mean scores made by the pre-college group on these eleven motivation and attitude tests are given in Table 50. They may be compared with the scores of the experimental and control groups shown in Tables 41 to 49 inclusive. To facilitate these three-way comparisons Tables 51 and 52 have been constructed. These tables show the relative positions for each of the experiments of the three groups; experimental, control and pre-college. If a group is significantly higher than one or both of the other groups, a capital H is shown in the appropriate column. A capital L represents a significantly low score. Lower case h and l are used to represent high and low mean scores that are not statistically significant. Dashes represent intermediate positions where there are no significant differences.

If comparisons are made across both Table 51 and Table 52 the results for all four experiments may be considered together. Students apparently enter college with high regard for a foreign culture, since the pre-college students scored high on this instrument in all four experiments,

Table 49

Means, F-Ratios, and Frequencies for Psychometric Tests; Programmed Learning Experiment; June 1963

Test	Mean Control	 Programmed Learning	 Experimental Materials	F-Ratio	N
Phylophilia	82.8	82.2	80.1	1.28	216
Anomie	27.2	30.0	28.8	1.68	216
Hours Study	6.6	8.5	8.0	4.98*	214
Total Motivation	13.0	14.0	14.5	3.11*	215
Foreigners Foreign Country	89.4	85.5	93.2	4.36*	213
Foreigners This Country	87.2	85.9	90.9	1.72	211
My FL Teacher	100.4	95.1	100.1	1.72	199
Me Spkg. Eng.	91.4	88.6	94.4	3.17*	197
Me Spkg. F.L.	70.9	73.6	77.5	1.84	196
Diff. Score	31.0	28.3	24.8	2.10	191
Seman. Sens.	20.3	19.4	22.0	1.66	211

* Significant at .05 level

Table 50

Mean Scores and Total Frequency for the Pre-College Control Group on the Eleven Psychometric Tests; 1963

Test	Mean	N
Phylophilia	83.1	213
Anomie	29.0	214
Hours Study	7.4	130
Total Motivation	14.8	134
Foreigners Foreign Country	86.7	210
Foreigners This Country	81.8	207
My FL Teacher	98.7	142
Me Spkg. Eng.	89.4	212
Me Spkg. F.L.	77.9	138
Diff. Score	22.6	135
Seman. Sens.	19.6	96

Table 51

Relative Positions of Means by Group and Test;
Withholding Orthography and Monitoring Experiments; 1963

Test	Withholding Orthography Experimental	Withholding Orthography Control	Withholding Orthography Pre-College	Monitoring Experimental	Monitoring Control	Monitoring Pre-College
Phylophilia	—	l	h	L	l	H
Anomie	l	H	L	h	H	L
Hours Study	L	h	H	l	H	L
Total Motivation	L	L	H	L	h	H
Foreigners Foreign Country	l	—	h	l	L	H
Foreigners This Country	l	h	—	H	L	L
My FL teacher	h	l	—	l	L	H
Me Spkg. Eng.	—	l	h	l	h	—
Me Spkg. F.L.	—	l	h	L	L	H
Diff. Score	h	H	L	H	H	L
Seman. Sens.	h	H	L	h	l	l

H - Significantly higher score
L - Significantly lower score

h - higher score not significant
l - lower score not significant

Table 52

Relative Positions of Means by Group and Test;
Laboratory Type and Programmed Learning Experiments 1963

Test	Laboratory Type			Programmed Learning			
	Non-Recording	Recording	Pre-College	Programmed Learning	Experimental Materials	Control	Pre-College
Phylophilia	h	L	H	—	l	—	h
Anomie	h	l	—	h	—	l	—
Hours Study	H	l	L	H	l	l	L
Total Motivation	h	l	—	L	l	L	H
Foreigners Foreign Country	h	—	l	l	H	h	L
Foreigners This Country	h	—	l	h	H	H	L
My FL Teacher	h	—	l	l	—	h	—
Me Spkg. Eng.	—	h	l	l	H	h	L
Me Spkg. F.L.	l	—	h	l	l	L	H
Diff. Score	h	H	L	H	h	H	L
Seman. Sens.	l	h	—	l	—	h	—

H - Significantly higher score
L - Significantly lower score

h - higher score not significant
l - lower score not significant

although they were significantly higher only in the monitoring and laboratory type experiments. It is not surprising that in the two significant cases they were also low on anomie.

The scores of the pre-college group on hours of study are not comparable with the scores of the other two groups, in which the report was presumably an historical one, reporting the amount of time that had been spent studying language. The statement of the incoming freshmen was "the substance of things hoped for, the evidence of things not seen." The same comment applies in large part to the high scores of the pre-college group on the Total Motivation score, of which hours of study is one component.

On the attitudes towards foreigners in this country, the results are mixed. In two of the experiments there are no significant differences. In the monitoring experiment the pre-college group is high on foreigners in their own country but low on foreigners in this country. In the programmed learning experiment the pre-college group was significantly lower than the programmed learning group in one case and lower than both the programmed learning and experimental materials groups in the other. The importance of any statistically significant differences on the remaining attitude scales is doubtful. The pre-college students had not yet had sufficient experience with their foreign language teachers to provide a valid response on this question. It is, understandable, and perhaps reassuring, that they have a relatively low opinion of themselves speaking a foreign language as compared with their self-estimation as speakers of English. It is also understandable that they are low in the single significant comparison involving semantic sensitivity.

Reasons for Studying Language and Thinking in the Foreign Language

There were eight tests in the 19 taken or adapted from the Colorado study for which computation of means did not seem appropriate. These were two questionnaires about reasons for studying foreign language (Forms I-A and I-B shown in Appendix 4) and a questionnaire designed to elicit information on tendency to think in the foreign language. The reasons for this statistical treatment have been stated earlier in this chapter, (in the footnote page 225) but it may clarify the discussion to explain them in more detail as they relate to the instruments to which they apply.

In Form I-A a student was asked to rank the following statements in the order in which they applied to him.

I AM STUDYING FRENCH (GERMAN, SPANISH) BECAUSE:

______a. I think it will someday be useful in getting a good job.

______b. I think it will help me to better understand the French (German, Spanish) people and their way of life.

______c. It will allow me to meet and converse with more and varied people.

______d. A knowledge of two languages will make me a better educated person.

______e. Any other personal reason:________________________

In Form I-B the student was asked why the study of the foreign language could be important to him, indicating his position on a scale running from, "Not my feeling at all," on one end to "definitely my feeling," at the other. Illustrative statements are:

> "I think it will someday be useful in getting a job"
>
> "I need it in order to graduate from college"

> "I feel that no one is really educated unless he is fluent in the French (German, Spanish) language"

These statements were classified as "integrative" or "instrumental", representing opposite poles on an "orientation index" developed originally by Wallace Lambert. "An instrumental or extrinsic reason for undertaking the study of a discipline is external to the acquisition of the skill for its own sake; thus a student might enroll in a physics course for the instrumental reason that it satisfies a college requirement or that he expects to be more admired if he understands physics. A more integrative or intrinsic reason for enrolling in a physics course might be that the student is interested in the subject matter for its own sake and wishes to understand the material out of sheer interest in it." (Colorado Report, Mimeo Form page V-49)

The reason for not computing averages of the ranking given by the student to the statements on Form I-A or of the numerical values assigned to positions on the scales used in Form I-B, is that there is no reason to believe that equal differences in numerical values represent equal differences in fact; that the difference between a first and second ranking of the statements on Form I-A, for example, is the same as a difference between a third and fourth ranking.

With data of this kind the appropriate procedure is to compare the distributions of responses. The test of significance used is the non-parametric, Chi-square test. Its application may be illustrated by showing the distribution of responses on Form I-A by college (See Table 53). The distribution of the responses is into six positions ranging from "Extreme Integrative" responses at one end to "Extreme Instrumental" responses at the other. For each position three figures are shown. The first of these shows the number of students at the designated college that gave the response indicated. This number is referred to as the actual frequency. The second number is the expected frequency. It shows the number at the particular college that would have given this reply if the persons responding from that college had responded in exactly the same proportions as did the students from all the colleges. The third number is the difference, the actual frequency minus the

Table 53

Distribution of "Integrative - Instrumental" Motivation Scores, Test I-A

College	Extreme Integrative					Extreme Instrumental	Total
I	35, 35, 0	66, 56, 10	30, 44, -14	81, 82, - 1	33, 29, 4	58, 57, 1	303
II	15, 18, - 3	29, 30, - 1	26, 23, 3	42, 44, - 2	17, 16, 1	33, 31, 2	162
III	24, 28, - 4	31, 45, -15	27, 34, - 7	75, 66, 9	29, 23, 6	55, 45, 10	241
IV	82, 73, 9	117, 118, -1	105, 91, 14	158, 174, -16	49, 62, -13	126, 119, 7	637
V	23, 34, -11	68, 56, 12	53, 43, 10	88, 82, 6	31, 29, 2	37, 56, -19	300
VI	23, 14, 9	16, 23, - 7	12, 17, - 5	37, 33, 4	12, 12, 0	22, 23, - 1	122

---

x^2 = 52.31, d.f. = 25, $p < .01$

expected frequency. It shows the extent to which the students at the colleges responded differentially to the questions. The students at College III, for example, made fewer than the expected responses toward the integrative end of the scale, and more than the expected number at the instrumental side. College IV has more than the expected number at the extreme instrumental end of the scale, but otherwise leans toward integrative responses. College V avoids the extremes, but aside from this tends to be more integrative than instrumental.

The test for significance in this type of analysis applies to the whole distribution. It relies on an estimate of the probability that differences as large as those observed between the actual and expected frequencies could arise by chance. It utilizes Chi-square, a total that increases with greater departures of actual from expected, and degrees of freedom, a number that depends on the number of cells in the table.

Table 53 was based on Form I-A, "Lamberts' Short Form." Table 54 is concerned with the same issues, but records

Table 54

Distribution of "Integrative - Instrumental" Motivation Scores, Test I-B*

College	Number of Scores, 5 or less	Number of Scores of 6	Number of Scores of 7	Number of Scores of 8	Number of Scores of 9	Number of Scores, 10 or more	Total
I	22, 31, - 9	6, 13,- 7	10, 13, - 3	12, 12, 0	18, 13, 5	47, 33, 14	115
II	61, 81, -20	38, 34, 4	28, 32, - 4	41, 31, 10	31, 33, - 2	97, 85, 12	296
III	67, 63, 4	28, 26, 2	19, 25, - 6	27, 24, 3	17, 26, - 9	73, 67, 6	231
IV	124, 82, 42	36, 34, 2	29, 32, - 3	30, 32, -2	25, 34, - 9	56, 86, -30	300
V	29, 43, -14	18, 18, 0	21, 17, 4	18, 17, 1	20, 18, 2	52, 45, 7	158
VI	168,171, - 3	70, 71,- 1	79, 67, 12	53, 66, -13	85, 71, 14	171,180, - 9	626

$X^2 = 81.68$ d.f. = 25, $p < .01$

* The more instrumental the response the higher the assigned score.

responses to "Lamberts' Long Form." (I-B) It tells, however, a somewhat different story. College I, which in Table 53 departed from the expected number primarily because of unbalance between adjoining cells on the integrative side (an imbalance in the direction of integrative responses), exhibits a tendency toward instrumental responses on Table 54. College II, which gave very closely the expected responses on Form I-A, leans in the direction of instrumental responses. The College III students are no longer so clearly instrumental in their responses. College IV shows a complete reversal of form. On Form I-A it had more than the expected numbers at the extreme end of the scales, but otherwise leaned toward instrumental responses. On Form I-B the responses are very heavily weighted in favor of extremely integrative responses and away from the extemely instrumental.

In view of these contrasts one wonders whether these two instruments are measuring the same thing. In each case the results are statistically significant, but it is not possible to say what the interpretation should be in terms of the types of motivation for language study.

The other series of responses to which the Chi-square test was applied were those on Form T which was concerned with ability to think in the foreign language: Five questions were scored separately and a total score was derived, based on the responses to the five questions. The five questions were:

T-1 = When speaking French (German, Spanish) to what extent do you feel it is necessary to translate the material into English?

T-2 = How often do English words occure to you when you are talking French (German, Spanish)?

T-3 = How often do French (German, Spanish) words occur to you when you are talking English?

T-5 = Has French (German, Spanish) entered into your dreams?

T-7 = To what extent do you feel capable of thinking in French (German, Spanish)?

The responses were made in different ways to the different questions. One called for a checkmark on a scale running from "always" to "never". In two cases students were asked to select "most of the time", "sometimes", "rarely", and "never". Other phrasing was used for other questions. Numerical values were assigned to positions such that higher numerical values indicate greater ability to think in the language.

In Table 55 the students' responses were to the question: "When speaking the foreign language, to what extent do you feel it is necessary to translate the material into English?" The responses at four of the colleges do not depart greatly from the expected, but the students at College IV report translation to an unusual extent, while the students at College VI in a much less striking manner exhibit confidence in their ability to speak without translating.

The answers of students to the question. "How often do English words occur to you when you are talking the foreign language?" are consistent with the responses to the first question. They are shown in Table 56. The students of College IV, who are given to translation when speaking the foreign language, also are more likely than students at the other colleges to have English words occur to them when speaking the foreign language. The students at College VI, on the other hand, say they have this experience "rarely" in greater proportion than the students at other colleges.

Presumably thinking in the foreign language is reflected

Table 55 "Thinking in the Language"

Responses to Question: "When speaking the foreign language, to what extent do you feel it is necessary to translate the material into English?

College	RESPONSES						
	"Never"					"Always"	Total
I.	15, 10, 5	38, 35, 3	28, 31, -3	22, 21, 1	16, 22, -6	4, 4, 0	123
II	33, 25, 8	85, 85, 0	82, 77, 5	48, 52, -4	49, 55, -6	6, 9, -3	303
III	17, 19, -2	73, 68, 5	67, 60, 7	40, 41, -1	36, 44, -8	6, 7, -1	239
IV	12, 25, -13	58, 85, -27	75, 77, -2	57, 51, 6	83, 55, 28	17, 9, 8	302
V	15, 13, 2	50, 46, 4	34, 41, -7	26, 28, -1	32, 29, 3	5, 5, 0	162
VI	52, 52, 0	196,181, 15	163,163, 0	109,109, 0	105,116, -11	15, 19, -4	640

$\chi^2 = 46.37$, d.f. = 25, $p < .01$

Table 56 "Thinking in the Language"

Responses to Question: "How often do English words occur to you when you are talking the foreign language?"

College	RESPONSES				
	"Never"	"Rarely"	"Sometimes"	"Always" or "Most of the time"	Total
I	8, 6, 2	53, 51, 2	53, 55, -2	8, 10, -2	122
II	23,16, 7	120,125, -5	138,137, 1	22, 25, -3	303
III	12,13, -1	112,100, 12	101,109, -8	17, 20, -3	243
IV	12,16, -4	81,125, -44	172,137, 35	37, 24, 13	302
V	7, 9, -2	63, 67, -4	74, 73, 1	18, 13, 5	161
VI	30,33, -3	305,266, 39	264,290, -26	42, 52, -10	641

$\chi^2 = 51.70$, d.f. = 15, $p < .01$

in having the foreign words occur to you when you are talking English. If this is the case, the responses to this query are not entirely consistent with the responses to the earlier questions. The students of College IV, who reported an unusual number of translators in their midst, now reports an unusual number who have foreign words occur to them "always", or at least "most of the time." (See Table 57) The College VI students, however, are still running true to form, for in larger than the expected numbers they report that foreign words occur to them "sometimes".

Foreign language in one's dreams is a rather rare phenomenon for students at this stage of language learning, and the responses to this inquiry (shown in Table 58) do not depart significantly from a chance distribution.

The poll-takers finally came out with the sixty-four dollar question, "To what extent do you feel capable of thinking in the foreign language?" And Table 59, recording the responses, shows some surprising results. The students at College IV, who reported an unusual number of translators, has an unusual number who are confident of their ability to

Table 57 "Thinking in the Language"

Responses to Question: "How often do foreign language words occur to you when you are talking English?"

College	R E S P O N S E S				
	"Never"	"Rarely"	"Sometimes"	"Always" or "Most of the Time"	Total
I	5, 4, 1	25, 20, 5	85, 90, - 5	8, 9, -1	123
II	16, 9, 7	58, 50, 8	213,223, -10	16,21, -5	303
III	12, 7, 5	48, 40, 8	170,178, - 8	12,17, -5	242
IV	1, 9, -8	49, 50, -1	219,222, - 3	33,21, 12	302
V	5, 5, 0	19, 27, -8	121,119, 2	17,11, 6	162
VI	14,19, -5	93,106, -13	496,471, 25	38,45, -7	641

$X^2 = 42.28$, d.f. = 15, $p < .01$

Table 58 "Thinking in the Language

Responses to Question: "Has foreign language entered into your dreams?

College	RESPONSES				
	"Never"	"Rarely"	"Sometimes"	"Very Frequently"	Total
I	78, 82, - 4	28, 24, 4	14, 14, 0	3, 3, 0	123
II	205,202, 3	60, 58, 2	33, 35, - 2	5, 8, - 3	303
III	179,162, 17	37, 46, -9	24, 28, - 4	2, 6, - 4	242
IV	181,202, -21	66, 58, 8	41, 34, 7	13, 7, 6	301
V	104,108, - 4	35, 31, 4	16, 19, - 3	7, 4, 3	162
VI	437,428, 9	114,123, -9	75, 73, 2	14,16, -2	640

$x^2 = 24.33$, d.f. = 15, $p > .05$

Table 59 "Thinking in the Language"

Responses to Question: "To what extent do you feel capable of thinking in the foreign language?"

College	RESPONSES					
	Not at all	With great difficulty	Fairly readily	Easily	As Easily as English	Total
I	4, 6, -2	57, 43, 14	52, 56, - 4	8, 12, - 4	2, 6, - 4	123
II	16, 14, 2	125,107, 18	127,139, -12	22, 29, - 7	13, 14, - 1	303
III	14, 11, 3	96, 85, 11	111,111, 0	11, 23, -12	10, 12, - 2	242
IV	10, 13, -3	61,107, -46	153,138, 15	49, 29, 20	29, 15, 14	302
V	9, 7, -2	48, 57, - 9	77, 74, 3	16, 16, 0	12, 8, 4	162
VI	26, 28, -2	238,226, 12	292,294, -2	65, 62, 3	20, 31, -11	641

X^2 = 82.49, d.f. = 20, $p < .01$

think in the foreign language. Perhaps, when actually speaking they would rather not try to think in the foreign language. The students of College VI, who had fewer than the expected numbers of translators, now has more than the usual number who admit to great difficulty in thinking in the foreign language. The students of College II, who on earlier questions had run pretty much true to form, also has unusual numbers who confess that they think in the foreign language with great difficulty.

The composite scores, shown in Table 60, are somewhat harder to interpret than the component responses. College II on balance seems to lean toward low scores, and so does College III. College IV has come to rest triumphantly with an excess of the highest scores. College VI now averages out a little on the low side.

Without question the colleges responded differentially to the questions about thinking in the language. The meanings of their differential responses are not so clear. It is not apparent whether in using these tests investigators look upon "thinking in the language" as desirable in itself and

Table 60 "Thinking in the Language"

Distribution of Total Scores on the Five Questions Listed in Tables 55 - 59

College	Number of Scores, 5 or less	Number of Scores of 6	Number of Scores of 7	Number of Scores of 8	Number of Scores of 9	Number of Scores, 10 or more	Total
I	33, 28, 5	18, 15, 3	17, 18, - 1	15, 15, 0	15, 14, 1	25, 33, - 8	123
II	83, 69, 14	39, 36, 3	31, 44, -13	51, 37, 14	29, 35, - 6	69, 81, -12	302
III	66, 55, 11	33, 29, 4	34, 35, - 1	26, 30, - 4	30, 28, 2	53, 65, -12	242
IV	39, 69, -30	25, 36, -11	41, 44, - 3	27, 37, -10	37, 35, 2	133, 81, 52	302
V	33, 37, - 4	20, 20 0	25, 24, 1	23, 19, 4	16, 18, - 2	45, 44, 1	162
VI	150,146, 4	78, 77, 1	111, 94, 17	74, 78, - 4	77, 74, 3	151,172, -21	641

X^2 = 80.85, d.f. = 25, p <.01

to what extent it is assumed that such thinking will be accompanied by greater fluency in the foreign tongue. It may be of incidental interest to report that in the basic language skills as measured by the MLA Language Proficiency Tests College IV and VI were not far apart, but College IV had higher scores on a majority of the tests.

In the eight tables in which responses to Forms I-A, I-B, and T were compared between colleges seven were significant at the .01 level. When the responses of students in experimental and control groups were analyzed in a similar manner no differences significant at the .01 level were found and very few significant at the .05 level. Because the results of the application of this type of analysis to differences between experimental and control groups were not in general significant, the results are not presented in this report.

X CONCLUSIONS

The contract with the Office of Education and the supplementary funds made possible significant changes in foreign language instruction at the colleges that are members of the Associated Colleges of the Midwest. The preceding pages of this report indicate the issues with which language departments were concerned, the procedures that were followed and some of the results. Many important results were intangible, not susceptible of measurement and difficult to appraise even in subjective terms.

The overall purpose of this program was the "improvement of instruction in foreign languages and literatures at all levels, including the pre-service, undergraduate course of training for prospective foreign language teachers." Language instruction at all levels was undoubtedly affected by the program. The experiences of the participating college teachers in the Summer Workshops and in the other aspects of the program also must have had an important bearing on their work with students planning to become language teachers.

The period covered by this program was one of considerable increase in the number of students majoring in the language departments of member colleges. This growth pattern had begun before this project was started and it clearly reflects a number of factors, including increased interest in study in foreign countries. Table 1 shows the number of majors at each of the member colleges from 1958-59 to 1962-63 and the number of language majors studying abroad from 1957-58 to 1962-63. In 1958-59 there were 81 language majors in the ten colleges. By 1962-63 the number had grown to 183, an increase of 126 percent. During this same period total enrollments at the ten colleges had increased by almost 20 percent. The number of students abroad reached a peak of 50 in 1961-62. Almost one third of this total is accounted for by the 15 students from Beloit. Under a new program being inaugurated in 1964 increasing numbers of Beloit students will be studying abroad.

Some of the language majors who have included foreign study in their college programs will undoubtedly become language teachers and the study in the foreign country will

TABLE 61

ENROLLMENTS, LANGUAGE MAJORS, AND LANGUAGE MAJORS STUDYING ABROAD
1958-59 to 1962-63

Year		1957-58	1958 - 59		1959 - 60		1960 - 61		1961 - 62		1962 - 63	
College	1962 - 1963 Enrollment	Number Abroad	Major	Number Abroad	Major	Number Abroad	Major	Number Abroad	Major	Number Abroad	Major	Number Abroad
Beloit	1,039	(4)	23	(7)	24	(8)	20	(5)	22	(15)	28	(5)
Carleton	1,337	(1)	10	(3)	12	(3)	8	(4)	11	(3)	15	(2)
Coe	898	(0)	2	(0)	4	(0)	2	(0)	5	(0)	4	(0)
Cornell	863	(1)	10	(2)	14	(3)	25	(6)	34	(7)	39	(7)
Grinnell	1,144	*	6		7		5		5		10	
Knox	1,107	(1)	5	(2)	7	(2)	17	(3)	19	(4)	24	(2)
Lawrence	1,178	(1)	6	(1)	11	(2)	10	(2)	15	(4)	18	(2)
Monmouth	923	(1)	9	(3)	9	(2)	9	(3)	15	(5)	26	(1)
Ripon	851	(0)	1	(0)	3	(0)	11	(0)	9	(3)	9	(3)
St. Olaf	2,094	(0)	9	(1)	7	(0)	8	(3)	10	(9)	10	(9)
TOTAL	11,434	(9)	81	(19)	98	(20)	110	(26)	145	(50)	183	(31)

* Study abroad increased from 1 or 2 to 3 or 4 per year.

contribute to their effectiveness in this role.

The major change in language instruction to which this program contributed was a shift in the direction of greater emphasis on the audio-lingual approach to language study. Representatives of the language departments at the member colleges were asked about major shifts in emphasis in recent years. Seven of the respondents testified to shifts in the direction of greater emphasis on the audio-lingual skills; two representatives said that these skills were already a significant part of their goals, but that the program had contributed to the realization of these already existing aims. One representative felt that there had been no shift in emphasis as among the four basic language skills. As Chapter IV indicates, at all the colleges instruction emphasizing the audio-lingual approach was greatly facilitated by the installation of new or expanded language laboratories during the period covered by this program. A major aspect of the program was the attempt to develop insights and materials that would contribute to the more effective use of these facilities.

A shift in the direction of greater emphasis on audio-lingual skills almost of necessity starts with beginning language instruction. Consequently the initial emphasis was largely on the elementary level. This was inevitable at the outset and the three-year period proved to be too short to do as much at higher levels as was originally hoped. At all times, however, there was realization of the importance of the study of foreign languages and literatures at the advanced levels and a desire to provide the more advanced students with the best possible intellectual experiences.

The Summer Workshops were generally rated by participants as most valuable aspects of the program. Participants had an opportunity to discuss goals and procedures in language instruction with colleagues from other colleges and with visiting consultants. They designed classroom experiments to test ideas with respect to language instruction and they reviewed available teaching and testing materials, and developed additional materials of their own.

The experimental program was important, not through the production of a number of definitively established conclusions, but through the experience of attempting to find out what the perceptible effects of different pedagogical practices might be. The following tentative conclusions -- none of which is likely to be accepted by all of the participants -- emerged from this aspect of the program.

1. In elementary foreign language study electro-mechanical devices can be enormously useful in providing students opportunities for intensive audio-lingual experiences.

2. Inexpensive language laboratories and even take-home phonograph discs can provide the beginning language student with valuable practice of audio-lingual skills.

3. There is reason to believe that, at the beginning level many students make optimum progress by uninterrupted practice without monitoring or play-back.

4. The case for withholding orthography in the beginning course at the college level is not clear. In any event the period of withholding should be relatively short and the facilities and materials for audio-lingual activities should be ample.

5. Desirable combinations of classroom and laboratory time are yet to be discovered, but the programmed learning experience suggests that most current schedules make insufficient use of the laboratory.

6. Uniformly satisfactory methods of placing students who have had previous language instruction in college courses have not been developed. An unfortunately distant possibility is that with the assistance of programmed learning materials language instruction at

the college level can become more largely individual.

7. The language laboratory has potentialities not yet fully realized to contribute to the individualization of instruction. Programmed learning materials give promise of providing a means for realizing these potentialities.

8. There is great need for programmed learning materials adapted to age level, linguistic ability and stage of language study. Full realization of the possibilities of this approach cannot come until such materials are available and have been fully tested.

The aspect of the testing program that made use of some of the instruments used in the "psycholingusitic" experimentation at the University of Colorado confirmed the low correlations between semantic sensitivity scores and pre-tests, but did not confirm the modest positive correlations obtained at Colorado between semantic sensitivity and language achievement tests.

The current study also obtained modest confirming correlations between motivation scores and language achievement and in addition recorded a positive correlation between the MLAT and hours of study on language reported by students.

In general the ACM data did not reveal a consistent pattern contrasting experimental and control groups in correlations that related selected psychological variables on the one hand to the other psychological variables and to pre-test and post-test scores on the other.

Some interesting contrasts in mean scores on the various psychological tests were found between pre-college students, students in experimental sections and students in control sections, but no consistent pattern of relationships among the three groups.

Generally differences between colleges on the psychological tests were more significant than differences between experimental and control groups. This was particularly true of the instruments reporting students' reasons for studying foreign languages and ability to think in a foreign language.

APPENDICES

The appendices are not included in this copy of the report. If you wish a copy of one or more of the appendices, please write to The Associated Colleges of the Midwest, 20 North Wacker Drive, Chicago 6, indicating the appendix in which you have an interest. A list of the appendices is given in the table of contents.